Health Risk Assessment Resulting from $PM_{2.5}$ indoor Exposition in Xuanwei and Fuyuan, China

September 2016 - May 2017

Student Pedro Miguel Gonçalves Leite.

University Institute for Risk Assessment Sciences, Utrecht University & Faculdade de Ciências da Universidade do Porto.

Coordinators Dr. George Downward & Joaquim C. G. Esteves da Silva.

Acknowledgments

First, I would like to thank my thesis coordinator during my internship abroad, George Downward of the Institute for Risk Assessment Sciences (IRAS), in Utrecht, the Netherlands, and to my thesis coordinator in Portugal, Joaquim Esteves of Faculdade de Ciências da Universidade do Porto. The door of both offices were always open whenever I ran into trouble or had some sort of question about my work. They always allowed me to be on my own, but steered me in the right direction every time I felt lost.

I would also like to thank the help from my colleges in IRAS for all the support and warm reception that I received. Without that support, my experience abroad could not have been the same and this successful.

Finally, I must express my very profound gratitude to my parents, my girlfriend that always stayed by my side and a special gratitude to my long-date friend João Amorim for providing me with unfailing support and continuous encouragement throughout the years, and through the process of writing this thesis. This accomplishment would not have been possible without all of them by my side. My sincere and big thank to all!

Preface

The Institute for Risk Assessment Sciences (IRAS), in Utrecht, the Netherlands, is an interfaculty research institute within the faculties of Veterinary Medicine, Medicine and Sciences of Utrecht University. IRAS provides education and research on the human health risks of exposure to potentially harmful agents in the environment, at the workplace and through the food chain. Effects on ecosystems are also considered.

A part of completing my Master's degree was a 5-month internship. Since I wanted to benefit from the experience of working and learning in another country, improve my English skills, and develop my knowledge of other cultures, I requested an Erasmus+ internship at IRAS.

Dr. George Downward agreed to be my mentor and include me in the research he was conducting about the effects of household air pollution from the use of solid fuels amongst the residents of Fuyuan and Xuanwei counties, China.

In this internship report, I will describe my experiences during my internship period. This internship report contains an overview of what I have learned, tasks and projects that I have worked on during my internship. While writing this report, I will also address new methods that I have learned during my internship and their applications.

Abstract

This internship research was divided into two main components - educational and analytical. In the educational element, the consolidation of skills in epidemiological analysis (including linear regression and mixed effects models) were used to reproduce the previous epidemiological findings of Dr. George Downward's work. In the analytical element, this new knowledge was applied in an investigation among non-smoking women in Xuanwei and Fuyuan, China, of the relationship between fuels use and lung function measurements.

Linear regression and linear mixed effects models were used to test the differences in $PM_{2.5}$ (particulate matter sized of, generally, 2.5 micrometers and smaller) exposure between stove and fuel combinations and to investigate which variables contributed to personal $PM_{2.5}$ exposure, respectively. The amount of $PM_{2.5}$ exposure for each combination was calculated and values were found to be significantly reduced if the individuals changed the type of combination (the lowest combination reported was smokeless coal and portable stove). Spirometry parameters were predicted for each individual and for each combination of stove and fuel was calculated and compared with the real values. A stepwise linear regression was used to investigate which variables of the study had more impact in each parameter of the breathing ratio and itself. A linear discriminant analysis was conducted to identify which variables of the study had higher discriminatory capability in the breathing ratio. The results showed that the combination with the higher $PM_{2.5}$ exposure was 352 $\mu g/m^3$. After an improvement in the stove and/or fuel used, the exposure levels could drop more than 100 $\mu g/m^3$ in some combinations. Even though the $PM_{2.5}$ exposure values were extremely high, only 3.03% of the population presented moderate chronic obstructive pulmonary disease (COPD). The results of this study showed that the variable that had the most impact in the breathing ratio was the body mass index (BMI) and that there was a significant benefit in the use of smokeless coal, when compared to smoky coal or wood. However, smokeless coal might also present other harmful effects similar to the ones caused by smoky coal or wood that are not directly related to $PM_{2.5}$ levels.

In the future, and since the amount of available data was reduced and not ideal, further investigations should be done to support the findings of this work.

Keywords

$PM_{2.5}$; Air pollution; COPD; Human Health; Spirometry.

Table of Contents

List of Figures

List of Tables

List of Annexes

List of Abbreviations

AM – Arithmetic mean
BaP – Benzo[a]pyrene
COPD – Chronic Obstructive Pulmonary Disease
BMI – Body Mass Index
IAP – Indoor Air Pollution
FEV1 – Forced Expiration Volume in 1 second
FVC – Forced Vital Capacity
GM – Geometric mean
GSD – Geometric Standard Deviation
Lm – Linear Model
Ln – Natural Logarithm
PAH – Polycyclic Aromatic Hydrocarbon
$PM_{2.5}$ – Particulate Matter with diameter ≥ 2.5 micrometers
GLI – Global Lungs Initiative
GOLD – Global Initiative for Chronic Obstructive Lung Disease
WHO – World Health Organization

1. Introduction

Health problems have been consistently linked with air pollution in countries all over the world, regardless of population income or development status (Hong, 1996; Murray & Lopez, 1996[1]; Cohen et al., 2004; Smith, Mehta & Feuz, 2004). Historically, public health attention has focused mainly on the risk from air pollution resulting from outdoor sources (Hong, 1996; Murray & Lopez, 1996[2]) as evidence indicates that outdoor air pollutants could have significant effects on human health, even at low levels. Furthermore, industrial and vehicular industrial emissions in populated areas of the developing world are rising at alarming rates (World Health Organization, 2017).

Indoor air pollution (IAP) may, however, bring far greater health risks than outdoor air pollution, since indoor exposure levels of many dangerous and important pollutants exceed their exposure from outdoor sources (Smith, 1993). Although outdoor sources often dominate the majority of air pollution emissions, in many populations that still use solid fuels and unvented stoves, indoor exposures tend to be more dangerous for human health because they have higher concentrations in smaller areas and the individuals spend too much time in those contaminated spaces (Smith, 1993).

Solid fuels (wood, coal, animal feces, crop waste, etc.) are used by approximately 3 billion people around the world, mainly from low-to-middle income countries. They are used for daily chores such as cooking and heating, frequently using unventilated fire-pits or rudimentary stoves (World Health Organization, 2017). This leads to high levels of exposure to IAP, which is a major source of mortality and morbidity worldwide, causing up to 4 million deaths annually from multiple diseases including chronic obstructive pulmonary disease (COPD), pneumonia, and lung cancer (Gordon et al., 2014).

1.1. Indoor Air Pollution

The chemical, biological and physical contamination of indoor air may result in adverse health effects, especially in low-to-middle income countries where the main source of IAP comes from the smoke of solid fuels used in domestic chores. Smoke may contain carcinogens including suspended particulate matter sized of, generally, 2.5 micrometers and smaller ($PM_{2.5}$), nitrogen dioxide (NO_2), sulphur dioxide (SO_2), carbon monoxide (CO), formaldehyde and polycyclic aromatic hydrocarbons (PAHs) (Glossary of environment statistics, 1997). The associated risk attributable to exposure of carcinogens from IAP is approximately 17% of the annual premature lung cancer deaths. The risk is disproportionately higher for women due to their traditional role in food preparation (World Health Organization, 2017) and the time spent home doing domestic chores.

1.2. Lung Cancer

The term "*Lung Cancer*" is used to refer to a malignant tumor characterized by uncontrolled cell growth within lung tissue, more specifically the bronchi, bronchioles and alveoli (National Cancer Institute, 2017). According to the World Health Organization (WHO), in 2012 there were 14 million new cases of cancer, where more than 1.8 million were lung cancer. Furthermore, there were 8.2 million deaths caused by cancer, where approximately 1.59 million (19.4% of all deaths by cancer) were caused by lung cancer, which represents the highest mortality of all cancers (illustrative Figures in Appendix A) (Stewart, 2014). It is also one of the most aggressive human cancers, with a 5-year survival rate of between 10-15% (Howlader, Noone and Krapcho, 2017).

Although lung cancer cases are mainly caused by smoking (Kendzia et al., 2012), approximately 25% of all cases aren't attributable to tobacco, with the proportion of never-smokers developing lung cancer increasing over time (Parkin et al., 2005). Numerous risk factors have been identified to explain the occurrence of lung cancer among never-smokers, including environmental tobacco smoke exposure (refers to being exposed to someone else's cigarette, cigar or pipe smoke (Ccohs.ca, 2017), occupational exposure, IAP, outdoor pollution, prior diseases and genetic factors (Sun, Schiller and Gazdar, 2007; Toh et al., 2006; Subramanian and Govindan, 2007).

1.3. Chronic Obstructive Pulmonary Disease

COPD is a term used to describe progressive lung diseases including emphysema, chronic bronchitis, refractory (non-reversible) asthma and some forms of bronchiectasis that are characterized by increasing breathlessness. Over one-third of premature deaths from COPD in adults, in low-to-middle income countries, are due to exposure to IAP. Women exposed to high levels of indoor smoke from solid fuels are 2 times more likely to suffer from COPD than women who use cleaner fuels. Among men (who already have a heightened risk of COPD due to their higher rates of smoking), exposure to indoor smoke nearly doubles that risk (Copdfoundation.org, 2017; World Health Organization, 2017).

1.4. Stages of Chronic Obstructive Pulmonary Disease

Pulmonary function tests, called spirometry, are a method of assessing lung function by measuring the volume of air that an individual is able to expel from their lungs after a maximal inspiration. This test checks the amount (volume in Liters) of air and speed (airflow) that can be exhaled (Bellamy et al., 2005). Such measurements are used to diagnose COPD and its severity:

- The volume in a one-second forced exhalation is called the forced expiratory volume in one second (FEV1), measured in Liters.
- The total exhaled breath is called the forced vital capacity (FVC), also measured in Liters.
- In people with a normal lung function, FEV1 is approximately 70% of FVC (Cold et al., 2017).

A commonly used classification system to describe how severe COPD is called GOLD (Global Initiative for Chronic Obstructive Lung Disease) staging, where the stage will affect what treatment the person gets. The GOLD system bases the stage of COPD on (Cold et al., 2017):

- The symptoms;
- How many times a COPD had gotten worse;
- Any time the person had to stay in the hospital because of the COPD had gotten worse;
- Spirometry.

The GOLD classification for COPD is divided in 5 stages ranging from 0 to 4, as we can see in the Table 1 (adapted from Spirometry.guru, 2017) below:

***Table 1** GOLD classification for COPD.*

Stage	Characteristics
0: At risk	• Normal spirometry • Chronic symptoms (cough, sputum production) • GOLD 0 was introduced in the GOLD 2001 publication, but was no longer used in GOLD 2010
1: Mild COPD	• FEV1/FVC < 70% • FEV1 > or equal to 80% predicted • With or without chronic symptoms (cough, sputum production)
2: Moderate COPD	• FEV1/FVC < 70% • FEV1 between 50% and 80% predicted • With or without chronic symptoms (cough, sputum production)
3: Severe COPD	• FEV1/FVC < 70% • FEV1 between 30% and 50% predicted

	• With or without chronic symptoms (cough, sputum production)
4: Very Severe COPD	• FEV1/FVC < 70% • FEV1 < or equal to 30% predicted or FEV1 < 50% predicted plus chronic respiratory failure

The breathing ratio, FEV1/FVC, is illustrated in Figure 1 below:

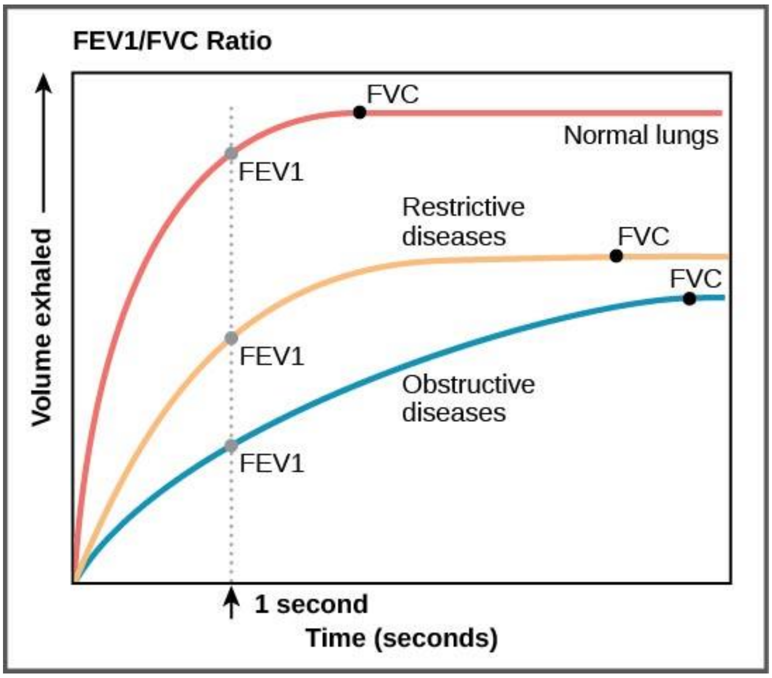


Figure 1 *Illustrative FEV1/FVC ratio graph used to diagnose whether a person has restrictive or obstructive lung disease (Boundless, 2016).*

1.5. Chinese Counties of Fuyuan and Xuanwei

Nowadays, half of the over one billion population of China still lives in rural environments (Tradingeconomics.com, 2017) where the use of solid fuels is still very frequent, as is the associated lung cancer risk (Enarson et al., 2009). The counties of Xuanwei and Fuyuan, located in North-East Yunnan province, have a population of approximately 2 million people. These are mostly rural areas, constituted by small villages, with the population living in poverty and where most resources come from farming. Their main source of energy for cooking and heating are solid fuels, coal being the most used, as there are still plenty of active coal mines.

From 1973 to 1975, a national cancer survey was performed by the government of China where it was reported that the annual age-adjusted rates for lung cancer mortality was 6.8 and 3.2 per 100.000 habitants for males and females, respectively.

The survey found that the lung cancer mortality rates in Yunnan province were lower than the national average for both sexes, 4.3 and 1.5 per 100.000 habitants, but in the Xuanwei county these rates were more than four times higher for men and much more for women, 27.7 and 25.3 per 100.000 habitants, respectively. Moreover, the county next to it, Fuyuan, had lung cancer rates of, approximately, more than a half as high as those found in Xuanwei. (Mumford et al., 1987), as we can see below in Figure 2 (Tian et al., 2008).

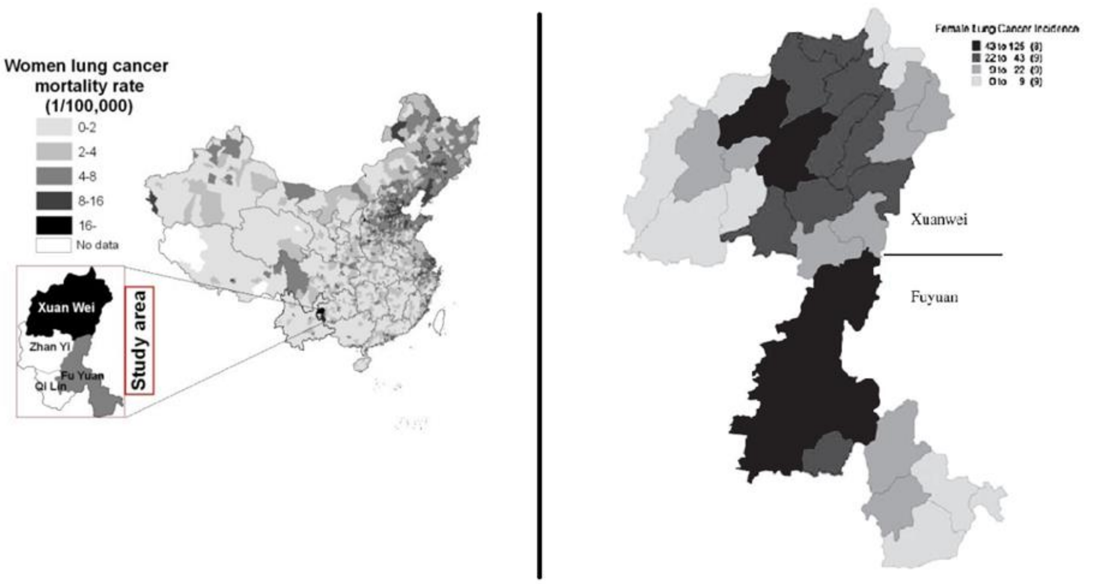


Figure 2 *On the left, is a map of China (not to scale) showing county-specific annual female lung cancer mortality rates in 1973-75. On the right, is a map of Xuanwei and Fuyuan counties (not to scale) highlighting geographic variation in lung cancer rates among females (adapted from Zhang, Lv and Sun, 2012).*

In Xuanwei and Fuyuan, like other rural areas all over China, solid fuels are routinely used for domestic chores, such as heating and cooking. The main solid fuel used is coal with a small proportion of the population using wood and other plant products. There are two widely used types of coal in the area, referred by locals as “Smoky” coal and “Smokeless” coal (bituminous and anthracite coal, see more in Appendix B). The names relate to the amount of smoke that each one of them emits during combustion. Previous epidemiological studies, first focusing in Xuanwei county, concluded that the use of smoky coal had a strong connection with the high lung cancer rates (Mumford et al., 1987; Mumford et al., 1989; Chapman et al., 1990). Smoky coal use was proportional to lung cancer mortality rates, as observed in Figure 3 below, where the villages with a higher percentage of smoky coal had higher lung cancer cases (Mumford et al., 1987).

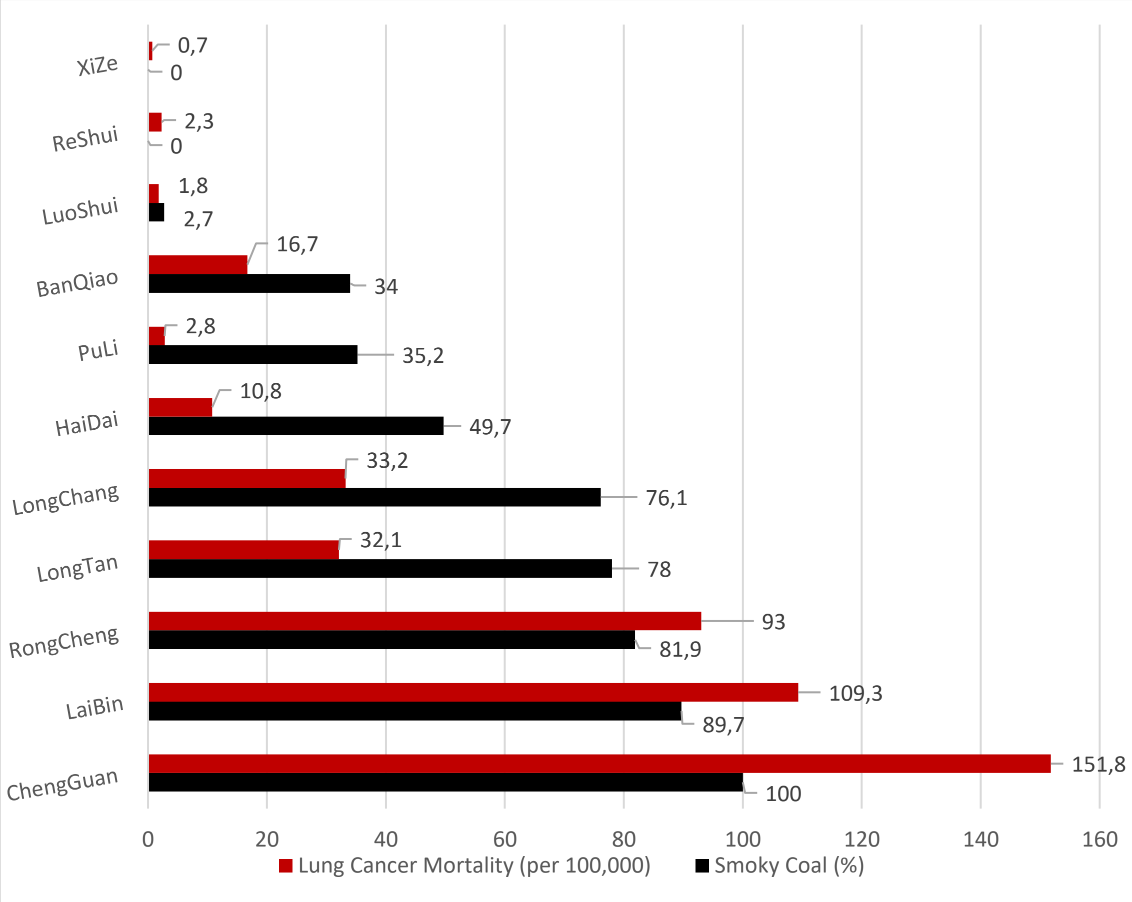


Figure 3 *Percentage of indoor smoky coal usage before 1958 and unadjusted lung cancer mortality in 1973-1975 in 11 Xuanwei villages (adapted from Mumford et. al. 1987).*

A case-control study, conducted from 1979 to 1983, investigating the etiology of lung cancer in the region found a weak association between smoking and lung cancer, but a strong association between domestic fuel types, suggesting that the effect of smoky coal on lung cancer is so strong that it over-rides the effect of smoking. A study performed by He et al. in 1991 showed that, in Xuanwei, more than 80% of men but less than 0.2% of women smoke tobacco, but the lung cancer and mortality rates in both sexes were similar, which makes it unlikely that tobacco smoking was the underlying cause, at least for women. Other risk factors identified were: the age that someone started cooking, the total number of years spent cooking and how many years of exposure to pollutants from the smoke of the solid fuels (He et al., 1991; Liang et al., 1988). After people started using ventilated stoves or switched to cleaner fuels the effect of smoking became more apparent (Kim et al., 2014).

Traditionally, people in Xuanwei and Fuyuan used solid fuels in unvented indoor fire-pits that would produce high levels of air pollution (Figure 4). After finding evidence of the link between smoky coal and lung cancer, many residents began the process of improving stoves from fire-pits to stoves with chimneys (vented stoves) (Lan et al., 2002). These improvements were made with two main purposes, first, to reduce the level of IAP, thereby reducing the risk of respiratory illness (Pandey et al., 1990; Naeher, Leaderer and Smith, 2000), and second, to reduce the demand for fuel by having a more efficient stove.

Figure 4 *Chinese woman cooking indoors over a traditional fire-pit with smoky coal in Xuanwei, China. A black circle was used to protect the identity of the person in the picture (Division of Cancer Epidemiology and Genetics - National Cancer Institute, 2017).*

Globally, improving stoves is a method used to reduce the amount of fuel used, improve burning efficiency and, most importantly, to reduce exposure to carcinogenic pollutants (Global Alliance for Clean Cook Stoves, 2017). The effects of stove improvements were studied in Xuanwei and Fuyuan and revealed that ten years after those improvements were put in practice, lung cancer rates had reduced over 50% (Lan et al., 2002). Also, research investigating the effects of using portable stoves, which are filled with coal, lit once outdoors and brought indoors after visible smoke has diminished substantially, thus reducing the exposure to pollutants, also showed a reduction in lung cancer rates (Hosgood et al., 2008). These findings indicate that exposure to carcinogenic pollutants present in the emissions of smoky coal can be reduced by improving the types of stoves and the type of ventilation (Hosgood et al., 2008).

The different villages use different types of coal, depending on the proximity of each available coal mine (1/3 coking coal, gas fat coal, coking coal and meager lean coal, smokeless coal and wood (more in Appendix B) (Figure 5). Smoky coal is the main coal type in the counties of Fuyuan and Xuanwei, which are both coal-rich areas, with numerous active mines still operating (Chen, 2000). A study performed by Downward in 2015, showed that the geographic location of smoky coal subtypes mines and the lung cancer rates in each county is positively related (Downward, 2015), meaning that villages near smoky coal mines presented higher lung cancer rates.

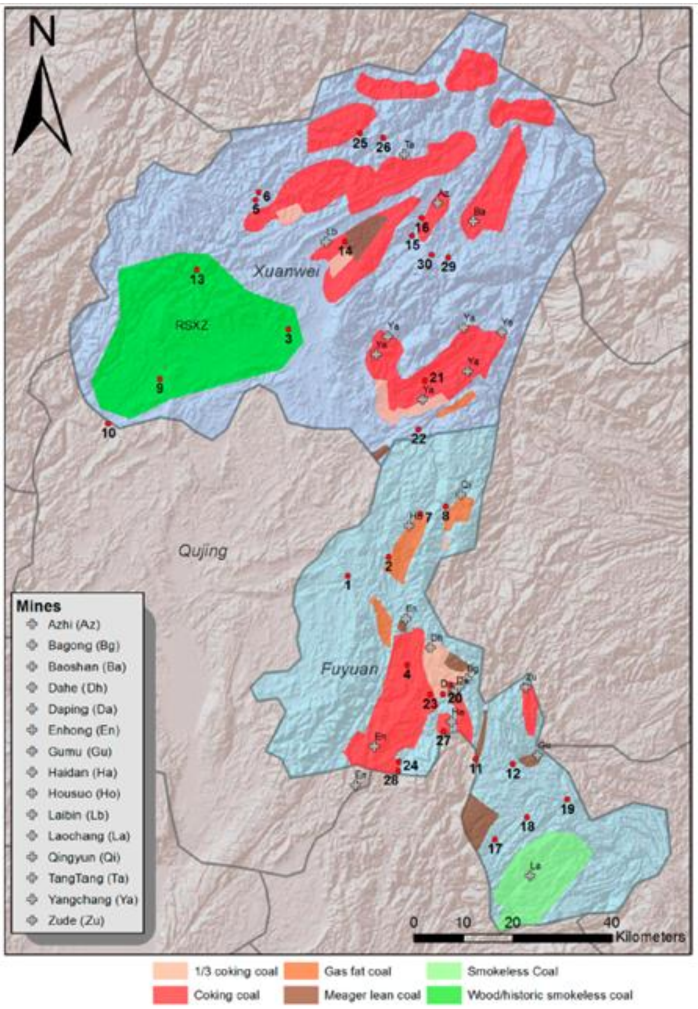


Figure 5 *Map of the counties of Fuyuan and Xuanwei. The location of the villages is represented by numbers as well as some of the mines reported in previous studies (Downward et al., 2014).*

1.6. Exposure to Pollutants from Solid Fuels

1.6.1. Particulate Matter 2.5

Particulate Matter with an aerodynamic diameter smaller than 2.5 micrometers (μm) is one of the principal pollutants in solid fuel smoke and has been associated with many adverse health effects. These particles are capable of carrying many toxic substances, passing through nose filtration, reaching the end of the respiratory track, penetrating deeply into the lung, irritating and corroding the alveolar wall, and, therefore, compromise lung function and even damage other parts of the body through air exchange in the lungs (Xing et al., 2016). Toxicology studies have shown that the combustion products from Xuanwei smoky coal are more carcinogenic and mutagenic than those from smokeless coal and wood products (Mumford et al., 1987; Liang et al., 1988). The burning of solid fuels generates very high indoor concentrations of airborne particulate matter (sometimes exceeding 20 mg/m^3, when the annual average, according to EPA legislation is 15 µg/m^3 (Kasteren and Konz, 2009)), PAHs and other organic compounds (Lan et al., 2008).

1.6.2. Polycyclic Aromatic Hydrocarbons

PAHs are a chemical group composed of more than one hundred organic compounds containing two or more condensed aromatic rings. They are produced by the incomplete combustion of organic material. Many of them, including benzo(a)pyrene (BaP) (which is commonly used subject of study as an indicator of total PAHs contamination), have been shown to be carcinogenic in experimental animals and regarded as potentially genotoxic (Caruso et al., 2008). According to the International Agency for Research on Cancer and Mutagenic, they are classified in the group 1, meaning that they are carcinogenic to humans as well (Rengarajan et al., 2015).

1.6.3. The Current Study

The research that will be described in this thesis was conducted in order to establish a better understanding about how solid fuel emissions, more precisely exposure to $PM_{2.5}$, might be linked to COPD on the regions of Xuanwei and Fuyuan, China. The specific goals of this thesis are as follows:

- Catalogue the exposures values of $PM_{2.5}$ for each combination of stove and fuel used;
- Explore the effect that changing the type of stove and fuel has in the exposures values of $PM_{2.5}$;
- Calculate the predicted spirometry values for each women of the study and compare with the real values;
- Explore what variables (including the role of stove and fuel) of the study might be having impact on spirometry values (FEV1, FVC, and FEV1/FVC ratio);
- Explore how the exposure values of $PM_{2.5}$ might be linked with spirometry values;
- Allow the findings in this thesis to be meaningfully applied to ongoing epidemiological research in Xuanwei and Fuyuan.

2. Materials and Methodology

This thesis focuses on the analysis of several variables that are part of the ongoing research of Dr. George Downward. The methods used to reach the results, firstly about the $PM_{2.5}$ exposure values and secondly about the spirometry values of each individual of the study, are described below, as well as the reasons why they were used.

2.1. Variables Under Investigation

Table 2 shows the abbreviation and corresponding codes of the most important variables used in this study. The other variables of this study can be found in appendix C.

Table 2 *Variable abbreviation table used in this thesis.*

Variable Abbreviation Table

AM	Arithmetic mean
AIC	Akaike Information Criterion
FEV1	Forced Expiration Volume in 1 second
FVC	Forced Vital Capacity
FEV1/FVC	Breathing Ratio
GM	Geometric mean
GSD	Geometric Standard Deviation
Lm	Linear Model
Ln	Natural Logarithm

2.2. Population Study

For this study, the population from previous research by Dr. George Downward was studied. This population consisted of only females, aged between 17 and 84, residing in Xuanwei and Fuyuan, who were primarily responsible for cooking and stayed longer inside. A total of 163 subjects were originally enrolled for study. Of those, 132 provided spirometry values of adequate quality for further study (see below) and were thus retained for analysis in this work. Those who weren't eligible were due to tests being considered with insufficient quality, particularly in regards with spirometry values. All study participants gave their written informed consent prior to being enrolled by the institutional review boards, US National Cancer Institute and China National Environmental Monitoring Centre (Downward, 2015). Spirometry values of each woman in the study (FVC, FEV1, FEV1/FVC) were selected because these are important indicators of lung function and used to identify respiratory problems.

2.3. Data Collection

2.3.1. Stove and Fuel Data Collection

Previous data was collected by Dr. George Downward, where each study participant completed a questionnaire, containing information relating to personal health, historical fuel and stove use, among others. During each measurement period, participants also reported all of their fuel and stove usage throughout their lives. Regarding what type of coal used, subjects directly reported whether they were using smoky or smokeless coal.

2.3.2. Particulate Matter Values Data Collection

Measurements of personal $PM_{2.5}$ inside the households were taken by drawing air through a 37mm Teflon filter mounted on a cyclone powered by a portable pump. The cyclone was attached near the breathing zone during the day and overnight the sampling arrangement was placed by the study participant's bed. Indoor measurements were also collected using the same equipment with devices placed between 1 to 2 meters from walls and stoves, as allowed by the size of the room. All potential pollution sources that may have contributed to outdoor air pollution (power plants, factories, etc.) within 5km of each village was documented, this information was gathered by asking inhabitants of each village (Downward, 2015).

2.3.3. Pulmonary Function Test’s Data Collection

Spirometry parameters such as FVC and FEV1, were collected from every subject using specific equipment and materials. An “EasyOne Spirometer”, respective spirette breathing tubes and the spirometry software “EasyWare” were used to calculate the FVC of each subject. The acceptance or exclusion criteria of the tests were based on the field protocol, which required that the subject was relaxed, did not wear tight clothing, performed all the procedure correctly, the area was checked for sharp edges in case subject faints during spirometry, the purpose of the test was explained, essential elements of the test were emphasized, and demonstrations of the procedure were done. The acceptability of the test depended on the cooperation of the participant and on the quality of the instructions of the physician that performed the tests, also the quality of the test provided was analyzed by a computer and an expert to ensure that a valid sample was provided. The reason why 31 out of 163 subjects weren't enrolled in the study was due to the quality rating of their test being insufficient.

2.4. Data Analysis Methodology

2.4.1. The Statistical Software

Statistical analysis was mainly performed using the program R, which is an open-source, free program that offers an environment for statistical computing and graphics (R-project.org, 2017).

2.4.2. $PM_{2.5}$ Exposure Data

Most of the data analyzed in Chapter 3 - "Previous information about the research subject" belongs to an educational and important step that allowed this thesis to be written, as a result of $PM_{2.5}$ data from previous Dr. George Downward studies.

2.4.3. Raw Data Analysis Methodology

Before advanced methods could be used, some raw data from previous studies needed to be analyzed and processed. For that reason, some basic calculations were made, such as arithmetic mean (AM), geometric mean (GM), the geometric standard deviation (GSD), as well as some histograms in order to understand the dispersion of the data.

2.4.3.1. Arithmetic Mean, Geometric Mean and Geometric Standard Deviation

The first calculations were the AM, GM and GSD of the multiple combinations of types of stoves and types of fuels, with the objective of calculating personal $PM_{2.5}$ ($\mu g/m^3$) exposure related to each combination. It was calculated using the **"***tapply***"** (Appendix D) formula from the program R, as showed below:

- tapply ($PM_{2.5}$ Data, list (Fuel data, Stove data), → Mean) / → Geomean) / → Geosd)

The second calculations were also the AM, GM and GSD, but this time to calculate the $PM_{2.5}$ exposures for each combination of coal type of each mine of both Xuanwei and Fuyuan.

- tapply ($PM_{2.5}$ Data, list (County, Smoky coal Subtypes data, Coal Mines data), → Mean) / → Geomean) / → Geosd)

2.4.3.2. Histogram

In order to use linear regression, the data must be normally distributed, for this reason, and before any calculation, whenever any analysis on data is done, it is important to see what kind of distribution the data has. Firstly, one histogram (Appendix D) was made with the data without any change and a second one using the natural logarithmic transformations. Natural logarithmic transformations of variables in a regression model are commonly used to handle situations where a non-linear relationship exists between the independent and dependent variables. Using the natural logarithm (ln) of one or more variables instead of the un-logged form makes the effective relationship non-linear, while still preserving the linear model. Natural logarithmic transformations are also a convenient mean of transforming a highly misrepresented variable into one that is more approximately normal (Benoit, 2011), both histograms were made with the following commands:

- hist($PM_{2.5}$ Data) ← First made
- hist(ln($PM_{2.5}$ Data) ← Second made after realizing that the data was not normally distributed

2.4.3.3. Linear Regression/Regression Analysis

In statistics, linear regression is an approach for modeling the relationship between a scalar dependent variable "y" and one or more explanatory variables (or independent variables) denoted "x" (Freedman, 2009). The linear regression equation is the following:

$$a = y + bx$$

- Where "a" stands for a constant term; it is the "y" intercept, the place where the line crosses the y-axis;
- Where "b" is the slope;
- Where "x" is the independent variable and "y" is the dependent variable;

Regression analysis is the statistical method used when both the response variable and the explanatory variable are usually continuous variables (i.e. real numbers with decimal places – used with heights, weights, volumes, or temperatures). Regression is the appropriate analysis when a scatterplot is the applicable graphic (in contrast to analysis of variance, when the plot would have been a box and whisker or a bar chart) (Crawley, 2012). In this study, the linear regression method was used to test differences in $PM_{2.5}$ exposure between differing stove and fuel configurations (Downward, 2015). The idea was to reach similar values, shown in Table 3, from previous studies made by Dr. George Downward. In Table 3, the "Ω" represents the values for the estimate linear effect modelling of Ln-Transformed personal $PM_{2.5}$ exposure for different fuel types, the "Ψ" represents the different stove designs and the "Φ" represents the reference value in μg/m³.

Table 3 *Linear mixed effect modelling of Ln-transformed personal $PM_{2.5}$ exposures (adapted from Downward, 2015).*

		Estimate
(Ω) Fuel Type	Smokeless Coal	Ref.
	Smoky Coal	0.27
	"Mixed" Coal	0.35
	Wood	1.03
	Plant Materials	0.43
	"Mixed" Fuel	0.37
(Ψ) Stove Design	Vented Stove	Ref.
	Unvented Stove	0.48
	Portable Stove	0.26
	Fire-pit	0.38
	Mixed Ventilation	0.2
	Unknown Ventilation	-0.34
(Φ) Reference Value*, in μg/m3		4.35

**Reference value represents base value of log transformed $PM_{2.5}$ in model for reference group (smokeless coal burnt in a vented stove, during autumn in a room with no windows).*

After applying the natural logarithmic transformation, in order to achieve a well distributed data, the linear model formula "lm" (Appendix C) was used to calculate the linear regression. The formula used was the following:

- Summary (linear model (ln (y) ~ ($x_1 + x_2 + \ldots + x_n$))
- Summary (lm (ln ($PM_{2.5}$ Data) ~ Fuel type data +Stove ventilation type)

2.4.3.4. Linear Mixed Effects Model

This model describes the relationship between a response variable and some covariates that have been measured or observed along with the response. In mixed effect models at least one of the covariates is a categorical covariate representing experimental or observational "units" in the data (A Simple Linear Mixed-effects Model, 2010). This model can be sorted in two categorical explanatory variables: the fixed effects, that influence only the mean of "y"; and the random effects, that influence only the variance of "y". While fixed effects are unknown constants to be estimated from the data and have informative factor levels, random effects govern the variance-covariance structure of the response variable, often have uninformative factor levels and have factors drawn from a large, sometimes very large, population in which the individuals differ in many different ways, but it isn't known exactly how or why they differ. Some examples are shown below (Table 4) to better explain the difference between fixed effects and random effects (Crawley, 2012):

Table 4 *Examples of fixed and random effects (adapted from Crawley, 2012).*

Fixed Effects	Random Effects
Drug administered or not	Genotype
Insecticide sprayed or not	Brood
Nutrient added or not	Block within a field
One country versus another	Split plot within a plot
Male or female	History of development
Upland or lowland	Household
Wet versus dry	Individuals with repeated measures
Light versus shade	Family
One age versus another	Parent

The linear mixed effects model was conducted to identify variables that contributed to personal $PM_{2.5}$ exposure. Like in the case of the linear regression model, the *"ln"* transformation was used in the formula in order to have well distributed values. The package "*lme4"* was used in the program R and the formula used was:

- model = lmer (ln (Y) ~ X_1 + … + X_n + (1|Random), REML=FALSE, data=data)

(REML= FALSE is used in case of comparing models with different "*Fixed Effects"* (during the simplification of the model), which is the case. The final formula used to get the results was:

- Mixed Effects Final Model = lmer (ln ($PM_{2.5}$ Data) ~ Fuel type data + Stove ventilation type + (1|Subject ID), REML=FALSE, data=data)

Appendix D can be consulted for more information about the formulas used.

2.5. New Scientific Findings - Analytical Part

After a complete literature review, analysis and study of the previous data, studies and information regarding $PM_{2.5}$ exposure, types of fuels and stoves used, it was acceptable to start new research using spirometry data.

2.5.1. Fuel and Stove $PM_{2.5}$ Exposure Combination

In order to understand which combinations of fuel and stoves was responsible for the highest exposure of $PM_{2.5}$, a predicted graphic was produced based in the data collected. This last step, regarding the analysis and processing of data about $PM_{2.5}$ exposure values, marks the beginning of the new scientific findings of this thesis.

2.5.2. Spirometry Data

Spirometry data was analyzed and processed considering all parameters of the pulmonary function test (FVC, FEV1 and FEV1/FVC) in order to fully correlate exposure to $PM_{2.5}$ with lung function and breathing problems.

2.5.3. Values for Predictive Spirometry - The Global Lung Function

The objective of this function is to establish international spirometry reference equations and values that are based on individual lung function data under standardized measurement conditions. They are modelled using modern statistical techniques, allowing the calculation of a predictive value for each spirometry parameter in a flexible and appropriate way where it's possible to adjust the equation for the heterogeneity of variability according to sex, ethnic group, age and lung function parameters. In this way, it is possible to compare real spirometry values with the predicted ones (Quanjer et al., 2012). The calculation of these predicted spirometry values was conducted using the Global Lung Function sheet calculator created by the Global Lungs Initiative (Webmaster, 2017).

2.5.4. Stepwise Regression Model

A stepwise regression model is a method of fitting various regression models, in which the choice of predictive variables is carried out by an automatic procedure (Hocking, 1976). In each step, a variable is considered for addition or subtraction from the set of explanatory variables based on some pre-specified criterion, in this study, based on the AIC. This method was conducted to identify variables that contributed to the variance of the breathing ratio were the final model chosen, was the one with the best AIC (the lowest value). In this model, the variable "y" was the breathing ratio (FEV1/FVC) and the co-variables "x" were all the other parameters gathered in the study (appendix C), except the individual parameters of the breathing ratio (FEV1 and FVC). They were both excluded since any variation on them will affect the breathing ratio since they are used in the calculation of the breathing ratio.

2.5.5. Linear Discriminant Analysis

Linear discriminant analysis (LDA) is a technique of data classification used when the within-class frequencies are unequal and their performances has been examined on randomly generated test data. This method allows to maximize the ratio of between-class variance to the within-class variance in any particular data set, guaranteeing maximal separability. It is used to determine which variable has higher contribution for the variance of discriminant function (Balakrishnama and Ganapathiraju, 2007). In this study, LDA was applied to identify which variables had bigger discriminatory power, in other words, impact, on the breathing ratio.

3. Previous Information About the Research Subject

In this chapter, the previous findings from Dr. George Downward's studies will be analyzed and used to cross with new data (Chapter 4 – "Results and Discussion"). This information was used to make an introduction to the values of each fuel type, stove type and how the $PM_{2.5}$ values were distributed.

3.1. First Look of Raw Data of Previous Studies

The objective was to reach the same results as Dr. George Downward got in his thesis in order to help continue his work.

3.1.1. Particulate Matter Screening Analysis

Table 5, shows the AM, GM and GSD for the personal $PM_{2.5}$ exposure related to all combinations of stove ventilation and fuel type.

Table 5 *Personal $PM_{2.5}$ (µg/m^3) exposure related to different stove ventilation configurations and fuel type (adapted from George Downward, 2015). N - number of observations, AM - Arithmetic Mean, GM - Geometric Mean and GSD - Geometric Standard Deviation.*

Fuel Type	Stove Design	N	AM	GM	GSD
Smoky Coal	Vented stove	110	150	134	1.6
	Unvented Stove	8	252	233	1.6
	Portable Stove	22	178	143	1.9
	Fire-pit	15	307	277	1.6
	Mixed Ventilation Stove	44	219	164	2.3
	Overall	**206*4**	**180**	**148**	**1.9**
Smokeless Coal	Vented Stove	5	151	126	2
	Unvented Stove	18	167	109	2.1
	Portable Stove	19	150	123	1.9
	Fire-pit	3	104	102	1.3
	Mixed Ventilation Stove	2	97	95	1.3
	Overall	**47**	**152**	**115**	**1.9**
"Mixed" Coal *1	Vented Stove	13	152	137	1.7
	Unvented Stove	0	-	-	-
	Portable Stove	14	209	180	1.8
	Fire-pit	2	156	150	1.5

	Mixed Ventilation Stove	9	192	176	1.6
	Overall	**38**	**183**	**161**	**1.7**
Wood	Vented Stove	8	226	183	1.9
	Unvented Stove	0	-	-	-
	Portable Stove	6	327	320	1.3
	Fire-pit	10	508	392	2.4
	Mixed Ventilation Stove	0	-	-	-
	Overall	**24**	**369**	**289**	**2.1**
Plant Materials *2	Vented Stove	3	123	109	1.8
	Unvented Stove	3	416	408	1.3
	Portable Stove	2	439	439	1
	Fire-pit	1	146	138	1.5
	Mixed Ventilation Stove	1	605	605	-
	Overall	**13*4**	**284**	**225**	**2.1**
"Mixed" Fuel *3	Vented Stove	19	121	104	1.8
	Unvented Stove	17	306	250	2.2
	Portable Stove	7	219	203	1.5
	Fire-pit	0	-	-	-
	Mixed Ventilation Stove	47	207	165	1.9
	Overall	**94*4**	**205**	**160**	**2**

**1 Refers to the use of combinations of smoky, smokeless coal, and prepared coal briquettes.*
**2 Plant materials include combinations of wood, tobacco stem and corncob.*
**3 Refers to combinations of wood, plant materials and coal.*
**4 Data for unknown ventilation stove or unknown fuel type are not shown but included in the overall.*

Table 6, shows the AM, GM and GSD for the personal $PM_{2.5}$ concentrations of all sub-types of smoky coal in each County and coal mine.

Table 6 *Personal $PM_{2.5}$ ($\mu g/m^3$) concentrations from smoky coal burning homes from Xuanwei and Fuyuan, by coal source (adapted from George Downward, 2015). N - number of observations, AM - Arithmetic Mean, GM - Geometric Mean and GSD - Geometric Standard Deviation.*

County	Smoky Coal Subtype	Coal Mine	N	AM	GM	GSD
Xuanwei	Coking Coal	Azhi	34	227	181	1.9
		Baoshan	12	210	168	2.2
		Laibin	28	153	132	2.1
		Tangtang	31	194	152	2
		Yangchang	14	142	125	1.6
		Overall	**119**	**189**	**153**	**2**
Fuyuan	Coking Coal	Daping	9	111	104	1.5
		Enhong	9	241	208	1.8
		Haidan	5	348	329	1.4
	1/3 of coking	Bagong	10	207	194	1.4
		Dahe	3	104	96	1.6
	Gas Fat Coal	Housuo	38	130	116	1.6
		Qingyun	2	237	237	1
	Meager Lean Coal	Gumu	4	138	96	2.8
		Overall	**80**	**168**	**142**	**1.8**

Figure 6 shows the histogram of the $PM_{2.5}$ raw data without any natural logarithmic transformation, representing values that were not well distributed. A total of 422 observations of $PM_{2.5}$ were made. Some individuals and household were sampled multiple times and in different temporal spaces. Measurements were made from August 28th 2008 to June 21st 2009.

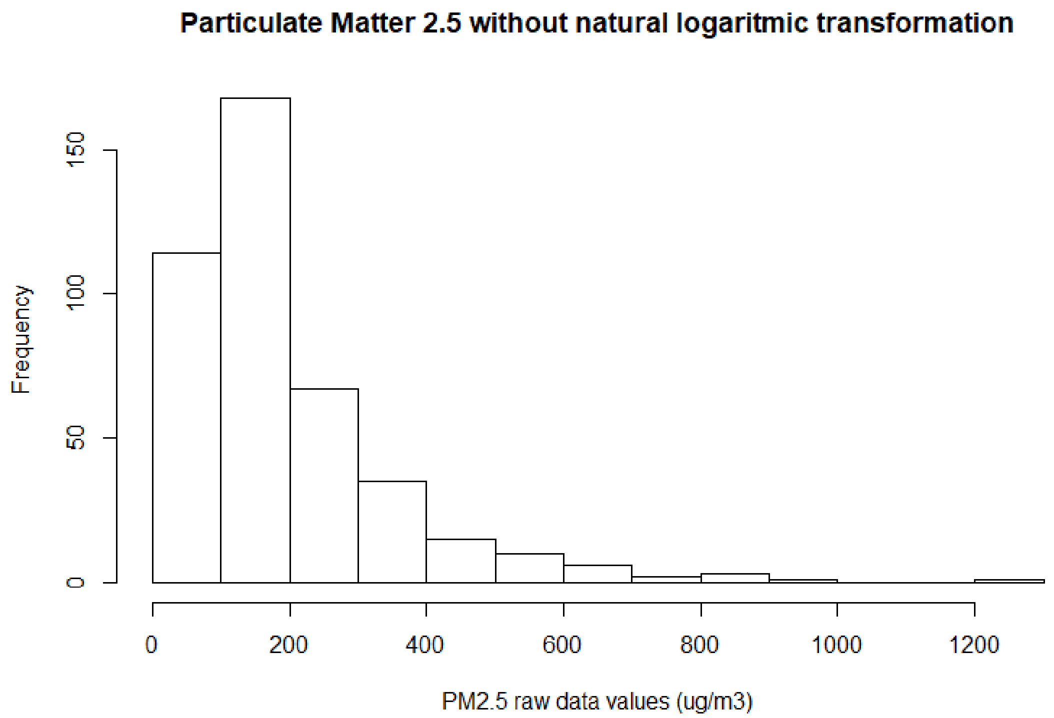


Figure 6 *Raw data from $PM_{2.5}$ exposure calculated without natural logarithmic transformation. Frequency represents the number of observations made.*

Since the data was not well distributed, the natural logarithmic transformations method was used, the results are represented in Figure 7.

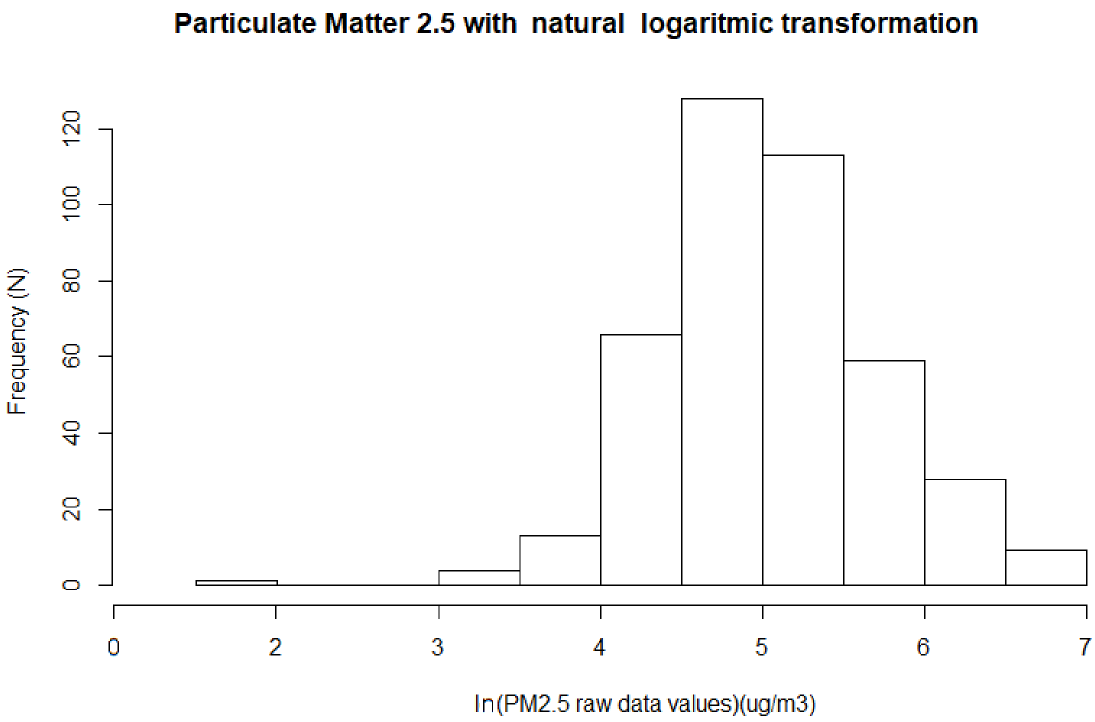


Figure 7 *Raw data from $PM_{2.5}$ exposure calculated with natural logarithmic transformation. Frequency represents the number of cases in each range of values.*

Figure 7 is a histogram representation of the natural logarithmic transformation of the raw $PM_{2.5}$ data presented above in Table 5. With this transformation it was possible to apply the linear regression and linear mixed effects model to the data.

The linear regression model showed in Table 7 represents the dependent variable, the natural logarithmic transformations of $PM_{2.5}$ data, and the independent variables, the fuel and stove type data.

Table 7 *Results obtained from linear model of natural logarithmic transformations of $PM_{2.5}$ data and fuel and stove type data.*

Formula

lm (ln ($PM_{2.5}$ Data) ~ Fuel type data + Stove ventilation type

Residuals

Min	First Quadril	Median	Third Quadril	Max
-3.3778	-0.3806	-0.0253	0.4057	2.2475

Coefficients	Estimate	Std. Error	t value	Pr(>|t|)
(Φ) (Intercept is Smokeless Coal and Vented Stove)	4.36141	0.11574	37.683	< 2e-16 ***
(Ω) Types of Fuel				
Smoky Coal	0.50458	0.11298	4.466	1.03e-05 ***
Other Coals	0.51261	0.14234	3.601	0.000356 ***
Wood	1.00145	0.16711	5.993	4.52e-09 ***
Plant	0.69538	0.19868	3.5	0.000516 ***
Other Fuels	0.48792	0.12113	4.028	6.70e-05 ***
(Ψ) Types of Stove Ventilation				
Unvented	0.53862	0.11573	4.654	4.40e-06 ***
Portable Stove	0.32741	0.09543	3.431	0.000663 ***
Fire-pit	0.52984	0.1248	4.246	2.70e-05 ***
Mixed	0.25978	0.08359	3.108	0.002017 **
Unknown	-0.58053	0.19516	-2.975	0.003106 **

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.*
Residual standard error: 0.622 on 411 degrees of freedom.
Multiple R-squared: 0.1852.
Adjusted R-squared: 0.1654.
p-value: 5.227e-14.

In Table 7, it is important to note that: "Φ" represents the intercept value (the intercept value is the expected mean value of "y" when all "x"=0; "Ω" represents the correlation between the $PM_{2.5}$ calculated with the type of stove used by each random subject ID; and "Ψ" represents the correlation between the $PM_{2.5}$ calculated with the type of fuel used by each random subject ID.

Table 8 shows the results obtained from the linear mixed effects model of the $PM_{2.5}$ data over the fuel and stove type.

Table 8 *Results obtained from linear mixed effects model of natural logarithmic transformations of $PM_{2.5}$ data and fuel and stove type data.*

AIC (¥)	**BIC**	**logLik**	**deviance**	**df.resid**
789.1	841.6	-381.5	763.1	409
Scaled residuals:				
Min	**First Quadril**	**Median**	**Third Quadril**	**Max**
-5.6183	-0.5543	0.0097	0.5882	3.5731
Random effects:				
Groups	**Name**	**Variance**	**Std.Dev.**	
Subjects ID	(Intercept)	0.1037	0.3221	
Residual		0.2764	0.5257	
Number of observations: 422		**Groups:** 163 Subjects		
Fixed effects	**Estimate**	**Std. Error**	**t value**	
(Intercept is Smokeless coal and Vented Stove) (Φ)	4.47311	0.12956	34.52	
Unvented	0.45822	0.13011	3.52	
Portable Stove	0.29988	0.10775	2.78	
Firepit	0.4617	0.14131	3.27	
Mixed	0.24479	0.08957	2.73	
Unknown (Ω)	-0.46097	0.22253	-2.07	
Smoky Coal	0.40497	0.12584	3.22	
Other Coals	0.47876	0.15119	3.17	
Wood	0.93021	0.1844	5.04	
Plant	0.52163	0.20781	2.51	
Other Fuels (Ψ)	0.39299	0.13083	3	

In Table 8, it is important to note that: "Ψ" represents the coefficient values for the calculated linear effect modelling of ln-transformed personal $PM_{2.5}$ exposures for the different fuel types; "Ω" represents the different coefficients values of stove designs; "Φ" represents the reference value in ln(μg/m^3); and "¥" represents the AIC. When examining the variance values in the individual random effect, it should be close to 0 or even 0, with all the variance in the residual term. The variance in random factor reveals how much variability there is between individuals across all treatments, not the level of variance between individuals within each group (Gardiner, Luo and Roman, 2009).

Both the linear regression model and linear mixed effects model of natural logarithmic transformations of $PM_{2.5}$ data and fuel and stove type data (Tables 7 and 8) were necessary to perform the next step. A model with those values of each group (stove and fuel) and the natural logarithmic transformed $PM_{2.5}$ data was used to create a scatterplot in order to see if a positive regression was observed.

4. Results and Discussion

The following chapters are related with new scientific findings.

4.1. Analysis of the Fuel and Stove Types Facts

Figure 8 is a scatterplot that shows the correlation between the natural logarithmic transformed model of the $PM_{2.5}$ raw data (from Table 8, Chapter 3.1.1. - Particulate Matter Screening Analysis) and the natural logarithmic transformed model "stove+fuel". The "stove+fuel" transformation consisted of the creation of a natural logarithmic transformed linear mixed effect model where the independent variable "y" was the $PM_{2.5}$ raw data, the dependent variables "x" were the stove and fuel data and the random parameter were the studied individuals.

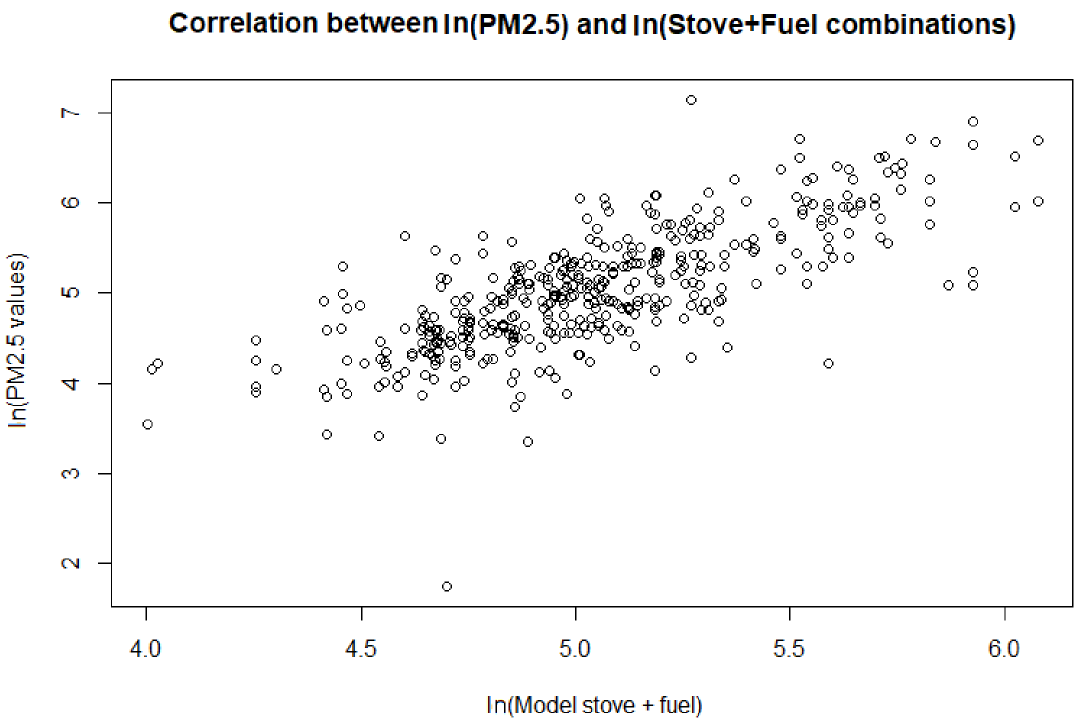


Figure 8 *Scatterplot of the correlation between ln-transformed $PM_{2.5}$ model and ln-transformed "stove+fuel" model.*

Figure 8 shows that the model had a positive regression, which indicated that the nature and strength of the relationship between "x" and "y" was positive as well. Since the $PM_{2.5}$ showed a positive correlation with the "stove+fuel" model, the next step was the calculation of all predicted $PM_{2.5}$ exposure for each combination of the study. The calculation consisted in adding all possible stove, fuel and intercept combination values ("Estimate" values from Table 8, Chapter 3.1.1. - Particulate Matter Screening Analysis), this information is presented in Table 9.

***Table 9** Calculations and results for $PM_{2.5}$ raw and predictive exposure (μg/m³) based in the values from the previous linear mixed effects model shown in Table 8.*

Fuel	Fuel value	Type of Ventilation	Ventilation value	Intercept	Fuel Value + Ventilation Value + Intercept	$PM_{2.5}$ Raw	$PM_{2.5}$ Predicted
Smoky	0.40497	Vented	0*1	4.47311	0.40497+0+4.47311	4.87808	**131.3781755**
Smoky	0.40497	Unvented	0.45822	4.47311	0.40497+0.45822+4.47311	5.33630	**207.7426388**
Smoky	0.40497	Portable Stove	0.29988	4.47311	0.40497+0.29988+4.47311	5.17796	**177.3207075**
Smoky	0.40497	Fire-Pit	0.4617	4.47311	0.40497+0.4617+4.47311	5.33978	**208.4668425**
Smoky	0.40497	Mixed	0.24479	4.47311	0.40497+0.24479+4.47311	5.12287	**167.8163120**
Smokeless	0*2	Vented	0*1	4.47311	0+0+4.47311	4.47311	**87.62882532**
Smokeless	0*2	Unvented	0.45822	4.47311	0+0.45822+4.47311	4.93133	**138.5636795**
Smokeless	0*2	Portable Stove	0.29988	4.47311	0+0.29988+4.47311	4.77299	**118.2723481**
Smokeless	0*2	Fire-Pit	0.4617	4.47311	0+0.4617+4.47311	4.93481	**139.0467211**
Smokeless	0*2	Mixed	0.24479	4.47311	0+0.24479+4.47311	4.7179	**111.9329465**
Other Coal	0.47876	Vented	0*1	4.47311	0.47876+0+4.47311	4.95187	**141.4392081**
Other Coal	0.47876	Unvented	0.45822	4.47311	0.47876+0.45822+4.4731	5.41009	**223.6517154**
Other Coal	0.47876	Portable Stove	0.29988	4.47311	0.47876+0.29988+4.47311	5.25175	**190.9000514**
Other Coal	0.47876	Fire-Pit	0.4617	4.47311	0.47876+0.4617+4.47311	5.41357	**224.4313792**
Other Coal	0.47876	Mixed	0.24479	4.47311	0.47876+0.24479+4.47311	5.19666	**180.6678026**

***1** Ventilation value for "Vented" is always 0 as it was used as reference for all other stove types.*
***2** Fuel value for "Smokeless" is always 0 as it was used as reference for all other fuel types.*

Table 9 *Calculations and results for $PM_{2.5}$ raw and predictive exposure (μg/m³) based in the values from the previous linear mixed effects model shown in Table 8 (cont.).*

Wood	0.93021	Vented	0*1	4.47311	0.93021+0+4.47311	5.40332	**222.1427071**
Wood	0.93021	Unvented	0.45822	4.47311	0.93021+00.45822+4.47312	5.86154	**351.2646753**
Wood	0.93021	Portable Stove	0.29988	4.47311	0.93021+0.29988+4.47311	5.70320	**299.8253085**
Wood	0.93021	Fire-Pit	0.4617	4.47311	0.93021+0.4617+4.47311	5.86502	**352.4892058**
Wood	0.93021	Mixed	0.24479	4.47311	0.93021+0.24479+4.47311	5.64811	**283.7546624**
Plant	0.52163	Vented	0*1	4.47311	0.52163+0+4.47311	4.99474	**147.6345554**
Plant	0.52163	Unvented	0.45822	4.47311	0.52163+00.45822+4.47312	5.45296	**233.4481508**
Plant	0.52163	Portable Stove	0.29988	4.47311	0.52163+0.29988+4.47311	5.29462	**199.2618921**
Plant	0.52163	Fire-Pit	0.4617	4.47311	0.52163+0.4617+4.47311	5.45644	**234.2619655**
Plant	0.52163	Mixed	0.24479	4.47311	0.52163+0.24479+4.47311	5.23953	**188.5814483**
Other Fuel	0.39299	Vented	0*1	4.47311	0.39299+0+4.47311	4.86610	**129.8136551**
Other Fuel	0.39299	Unvented	0.45822	4.47311	0.39299+00.45822+4.47312	5.32432	**205.2687303**
Other Fuel	0.39299	Portable Stove	0.29988	4.47311	0.39299+0.29988+4.47311	5.16598	**175.2090794**
Other Fuel	0.39299	Fire-Pit	0.4617	4.47311	0.39299+0.4617+4.47311	5.32780	**205.9843098**
Other Fuel	0.39299	Mixed	0.24479	4.47311	0.39299+0.24479+4.47311	5.11089	**165.8178671**

****1** Ventilation value for "Vented" is always 0 as it was used as reference for all other stove types.*
****2** Fuel value for "Smokeless" is always 0 as it was used as reference for all other fuel types.*

Since the fuel, stove and intercept values (Table 9) were obtained in a natural logarithmic transformation (from Table 8, Chapter 3.1.1. - Particulate Matter Screening Analysis), it was necessary to do a napierian exponential of the $PM_{2.5}$ raw data values to obtain the predicted $PM_{2.5}$ exposure values. The formula used is showed below:

$$e^{\text{PM2.5 “Raw”}}$$

The predictive values of the $PM_{2.5}$ for each stove and fuel combination are showed in Table 9 (Column “$PM_{2.5}$ Predicted”) and in Figure 9 below. These values indicate the predicted $PM_{2.5}$ exposure values for each combination of stove and fuel that the individuals of the study were potentially subjected.

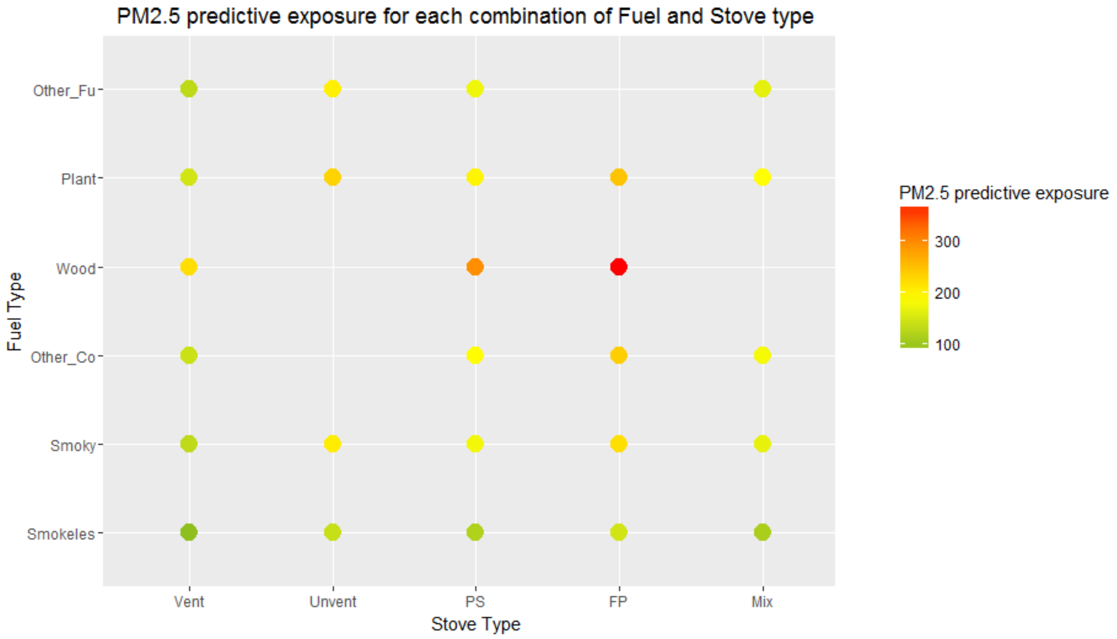
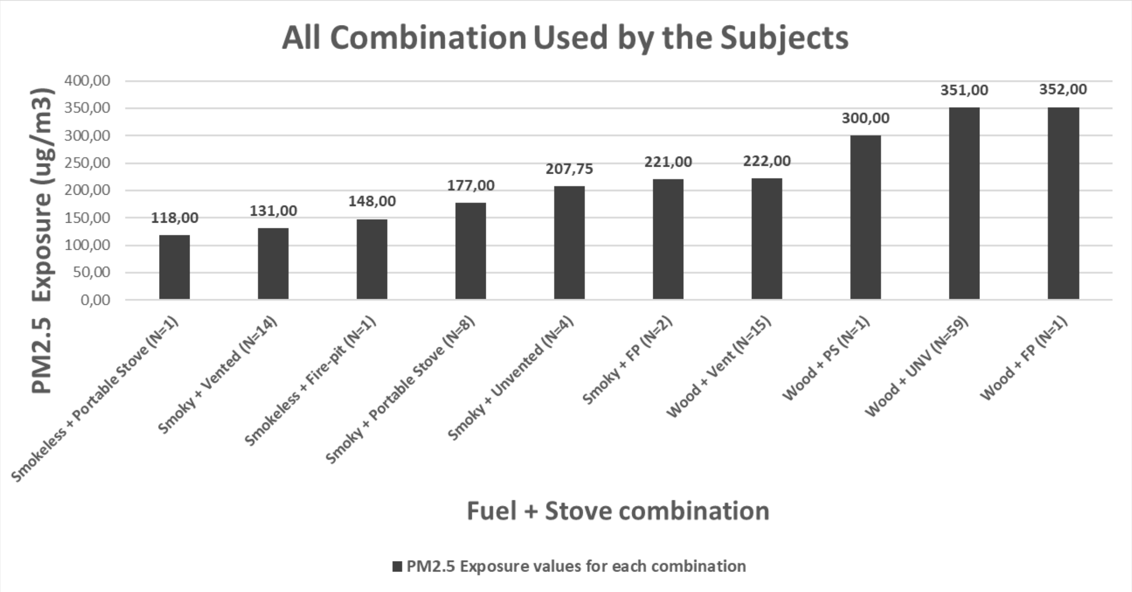


Figure 9 *$PM_{2.5}$ (µg/m^3) predictions for each fuel and stove combination.*

The populations of Xuanwei and Fuyuan counties are relatively poor and have multiple coal mines in their vicinity that are still active. For this reason, they primarily use fossil fuels and wood as fuel for multiple chores, such as cooking and warming, making themselves exposed to higher $PM_{2.5}$ concentrations (Downward et al., 2014). Figure 9 and Table 9 demonstrate that the use of a fire-pit with wood would be the combination with highest $PM_{2.5}$ exposure value, more precisely, 352.5 µg/m^3. The reported mean $PM_{2.5}$ values in Chinese cities was 61 µg/m^3, four times higher than the annual mean threshold value according to the EPA legislation - 15 µg/m^3. In fact, only 25 out of 190 big Chinese cities met the National Ambient Air Quality Standard (Zhang and Cao, 2015). For this reason, it was not surprising to see that the poor rural regions of Xuanwei and Fuyuan, where people still use rudimentary cooking and heating technology, and have overall poor living conditions, presented predicted $PM_{2.5}$ exposure values as high as 25 times the EPA recommendation. Even the lowest predicted $PM_{2.5}$ exposure value, 87.6 µg/m^3 for smokeless coal and vented stove, would still be almost 6 times higher than the recommended threshold.

Figure 10 below represents the fuel and stove combinations, as reported by the subjects of the study. A total of 106 out of 132 initial reports are represented. Unknown or missing reports were removed.



N *Number of reported subjects.*
Figure 10 *$PM_{2.5}$ exposure for each fuel and stove combination used by the subjects in the study.*

Figure 10 represents the values that each subject of the study was exposed when using a specific reported combination. Regarding the $PM_{2.5}$ values, the lowest reported combination belonged to "Smokeless + Portable Stove", 118 µg/m^3, with one observation. The highest reported combination belonged to "Wood + Fire-pit", 352 µg/m^3, with one observation. Regarding the type of fuel, the higher $PM_{2.5}$ exposure values were consistent with the amount of smoke and gaseous air pollutants release in their combustion. Wood had the highest $PM_{2.5}$ exposure levels and smoke production, followed by smoky coal (bituminous coal) and smokeless coal (anthracite coal) (Chafe et al., 2015). The difference between using a fire-pit and a vented stove, with wood as a fuel, resulted in a reduction of 130 µg/m^3 in $PM_{2.5}$ levels. Similarly, changing the type of fuel meant a difference of 204 µg/m^3, as in the case of wood to smokeless coal. This shows that the type of stove and/or fuel used in a household had an impact on the exposure to particles that are harmful to humans. Changing the type of fuel or creating a way to vent the area can, eventually, prevent respiratory diseases and reduce the lung cancer rate by reducing the exposure to $PM_{2.5}$.

4.2. The Example of Subject 372

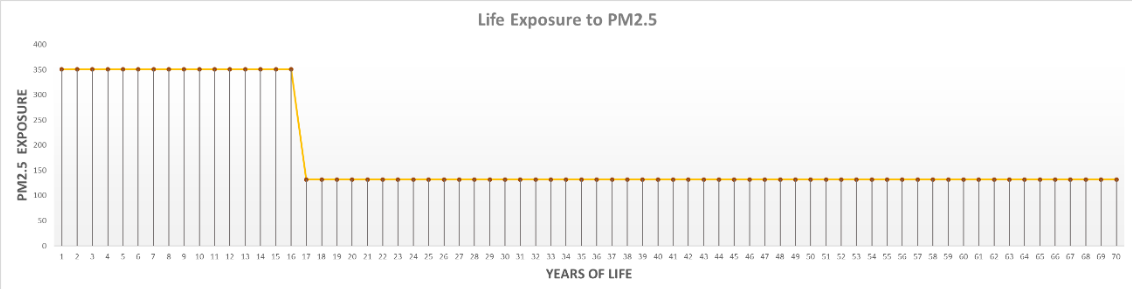


Figure 11 *Life exposure to $PM_{2.5}$ for the individual number 372.*

In Figure 11, we can see the variation in the predicted amount of $PM_{2.5}$ exposure during the lifetime of subject number 372, as reported by the individual. The subject was chosen to illustrate how the improvements in stove and fuel type can significantly affect the $PM_{2.5}$ values in a lifetime. Changing from an unvented stove and wood (351 µg/m^3) to a vented stove and smoky coal (131 µg/m^3), resulted in a reduction of 220 µg/m^3 in $PM_{2.5}$ levels after the age of 16.

4.3. Predictive Analysis of Raw Spirometry Data

Using the GLI calculator, it was possible to predict the spirometry values (FEV1, FVC and FEV1/FVC values) for each individual based on sex, age, ethnic group and height. These represent the spirometry values that each individual should have. Figure 12 below shows the difference between real and predicted values for breathing ratio (FEV1/FVC) for the 132 individuals of the study.

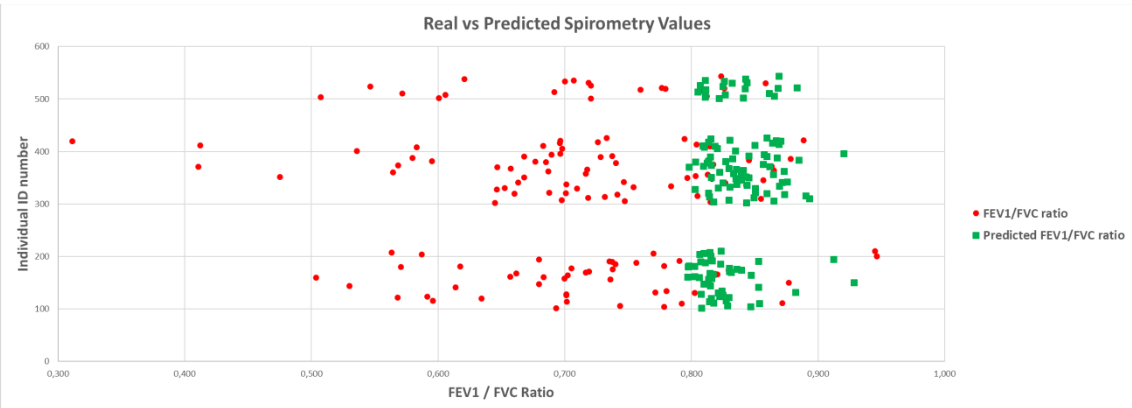


Figure 12 *Real vs predicted spirometry values for all individuals of the study.*

Real values, when compared with the predicted values, seem to be worse since the majority of the results were skewed below 0,7. It is important to remember that in people with normal lung function, FEV1 is approximately 70% of FVC (Cold et al., 2017). 54 of the 132 individuals presented real breathing ratio values below 0.7, contrasting with all predicted values very close or above 0.8. Moreover, one individual, subject 420, presented an alarmingly low breathing ratio value of 0.311.

The dispersion of the real and predicted values was calculated, the results are presented in Table 10.

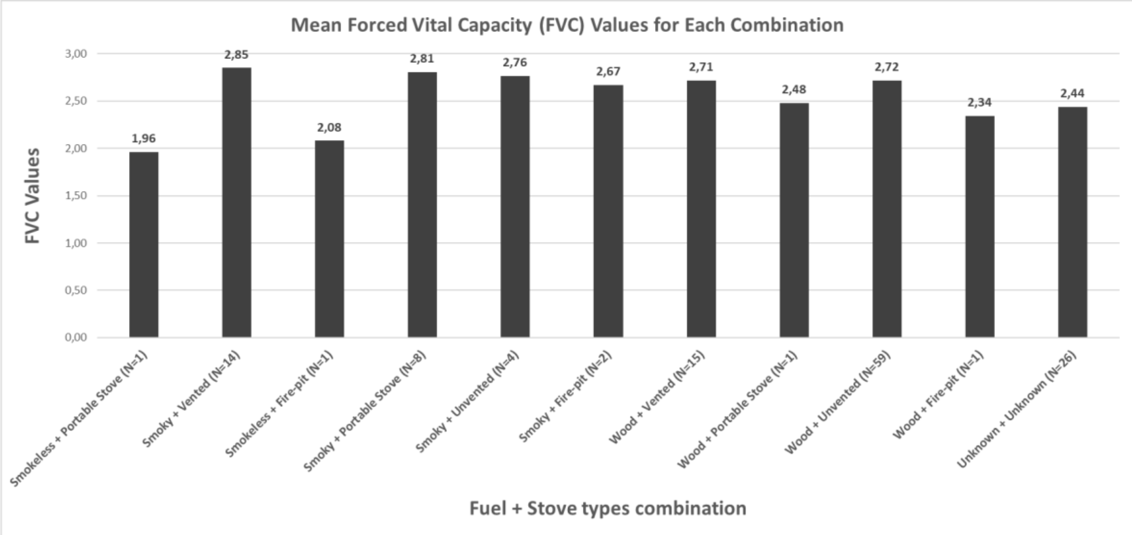
Table 10 *Standard deviation of the global real and predicted breathing ratio values.*

Real Values	~ 0.11
Predicted Values	~ 0.03

The dispersion values, showed in Table 10 and Figure 12, were relatively low in regards to the predictive results, but considerably higher when analyzing the real values. This variation is very common when studying natural biological systems. In particularly, humans show an intrinsically high diversity due to genetic, social and cultural features. The observation of a chronic environmental effect in the health of the studied populations is difficult, primarily because the variations that result from the exposure may be confounded with the intrinsic natural trends. Moreover, usually there is more than one environmental factor that contributes to a particular response. As a consequence, the variables that were analyzed to monitor the existence of a correlation between increased incidence of disease and environmental exposure present different sources of variation (inherent to human populations), that can explain the increased dispersion values. Consequently, mathematical/statistical techniques must be used to extract the required information.

4.3.1. Descriptive Spirometry Analysis

The mean FVC and mean FEV1 were calculated, for each combination of stove and fuel. Figure 13 and Table 11 showed that the worst mean FVC value (1.96) was associated with the combination “Smokeless + Portable Stove”, which is curious since it was the combination with the lower value of $PM_{2.5}$ exposure. However, this result may be due to the reduce amount of data, since only one individual reported the use of that combination.



N *Number of reported cases.*

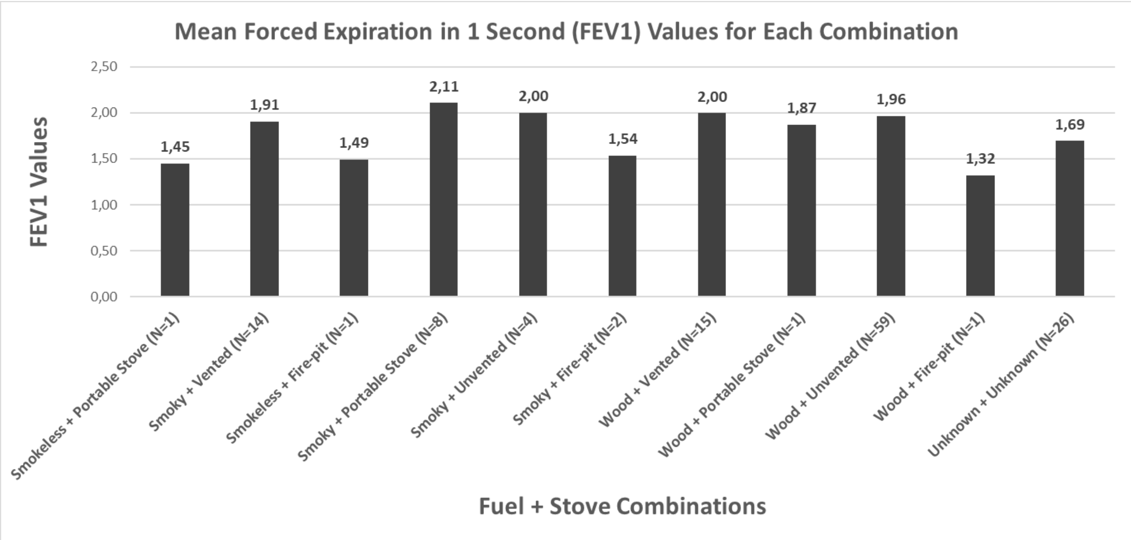
Figure 13 *Mean FVC values for each fuel and stove combination.*

Table 11 *Mean FVC and respective predicted values for each fuel and stove combination.*

Combinations	FVC	Predicted FVC
Smokeless + Portable Stove	1.96	1.96
Smoky + Vented	2.85	2.53
Smokeless + Fire-pit	2.08	2.25
Smoky + Portable Stove	2.81	2.64
Smoky + Unvented	2.76	2.13
Smoky + Fire-pit	2.67	2.11
Wood + Vented	2.71	2.50
Wood + Portable Stove	2.48	2.43
Wood + Unvented	2.72	2.35
Wood + Fire-pit	2.34	2.16
Unknown + Unknown	2.44	2.33

Surprisingly, only the combination “Smokeless + Fire-pit” had FVC values below the predicted results, meaning that the population had, in general, better lung capacity than what was expected. These results may indicate that the FVC is not the most affected parameter when considering $PM_{2.5}$ exposure.

Figure 14 and Table 12 showed that the worst mean value of FEV1 was associated with the combination "Smoky + Portable Stove", which is the 4th lowest value of $PM_{2.5}$ exposure.



N *Number of reported cases.*

Figure 14 *Mean FEV1 values for each fuel and stove combination.*

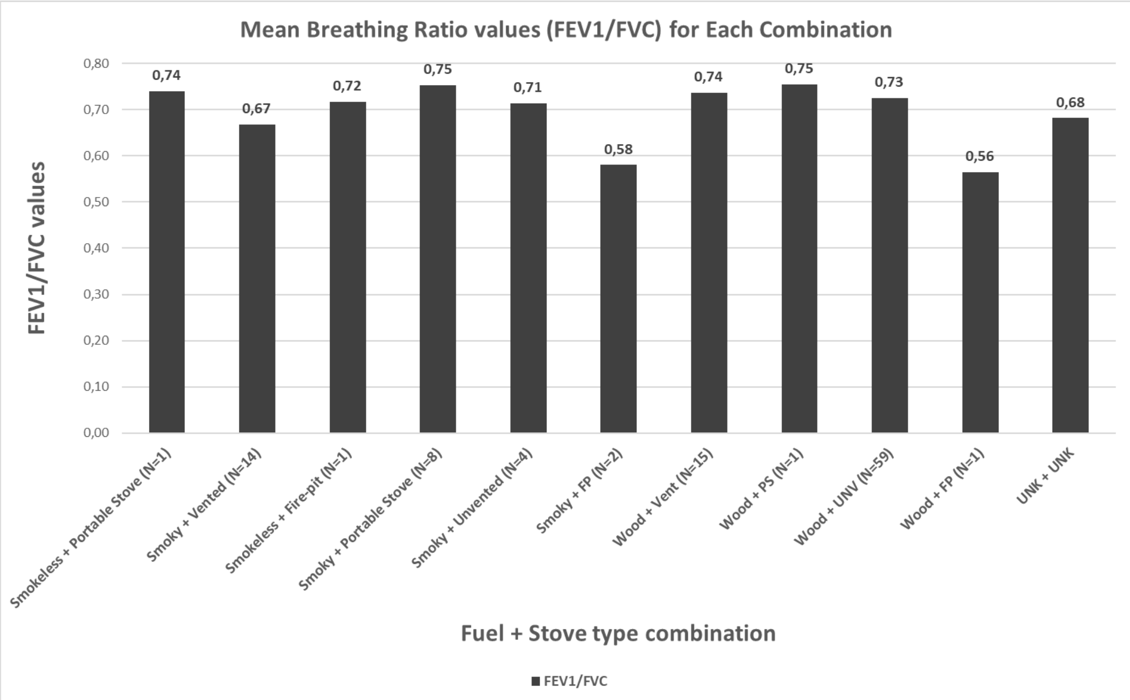
Table 12 *Mean FEV1 and respective predicted values for each fuel and stove combination.*

Combinations	FEV1	Predicted FEV1
Smokeless + Portable Stove	1.45	1.61
Smoky + Vented	1.91	2.14
Smokeless + Fire-pit	1.49	1.88
Smoky + Portable Stove	2.11	2.27
Smoky + Unvented	2.00	1.72
Smoky + Fire-pit	1.54	1.72
Wood + Vented	2.00	2.12
Wood + Portable Stove	1.87	2.02
Wood + Unvented	1.96	1.95
Wood + Fire-pit	1.32	1.77
Unknown + Unknown	1.69	1.94

Nine out of eleven combinations presented real FEV1 results lower than the predicted FEV1 values. Only two values of predicted FEV1 were inferior to the real FEV1 value: “Smoky + Unvented” and “Wood + Unvented” combinations.

4.3.2. Mean Breathing Ratio

In order to understand how the combinations might have affected the breathing ratio, the mean FEV1/FVC values were calculated for each combination and compared with the expected results (Figure 15 and Table 13).



N *Number of reported cases.*

Figure 15 *Mean breathing ratio values for each fuel and stove combination.*

Table 13 *Mean breathing ratio and respective predicted values for each fuel and stove combination.*

Combinations	FEV1/FVC	Predicted FEV1/FVC
Smokeless + Portable Stove	0.74	0.82
Smoky + Vented	0.67	0.85
Smokeless + Fire-pit	0.72	0.83
Smoky + Portable Stove	0.75	0.86
Smoky + Unvented	0.71	0.81
Smoky + Fire-pit	0.58	0.82
Wood + Vented	0.74	0.85
Wood + Portable Stove	0.75	0.83
Wood + Unvented	0.73	0.83
Wood + Fire-pit	0.56	0.82
Unknown + Unknown	0.68	0.83

As seen in Chapter 1.4 - "Stages of Chronic Obstructive Pulmonary Disease", the breathing ratio can be classified into five categories, when considering COPD risk: no risk (FEV1/FVC>0.7), mild (FEV1/FVC<0.7 and FEV1>0.8), moderate (FEV1/FVC<0.7 and 0.5<FEV1<0.8), severe (FEV1/FVC<0.7 and 0.3<FEV1<0.5) and very severe (FEV1/FVC<0.7 and FEV1<0.3). The results of Table 13 and Figure 15 showed that seven out of eleven reported combinations presented mean values of breathing ratio superior to 0.7, indicating no risk of COPD. However, the other four combinations are classified as presenting mild risk of COPD (FEV1/FVC<0.7 and FEV1>0.8). Particularly, the lowest mean value of breathing ratio calculated (0.56) was alarmingly low. This result was associated with the combination *"wood + fire-pit"*, which was the one with the highest $PM_{2.5}$ exposure value (352 µg/m^3). The second lowest mean value of breathing ratio calculated (0.58) was also alarmingly low. However, it was associated with the combination *"smoky coal + fire-pit"*, that was only the fifth highest $PM_{2.5}$ exposure value (221 µg/m^3). Even though these combinations seem to have affected the pulmonary function, higher $PM_{2.5}$ exposure levels were not directly related with the lowest breathing ratios. This result may indicate that these particles are part of the problem, but not the primary component in smoke responsible for negative effects on human health.

In China, COPD is becoming an important cause of public health concern and ranks first among the causes of disability (Murray and Lopez, 1996). Figure 16 indicates what COPD stage each one of the 132 individuals was in, according to the GOLD classification.

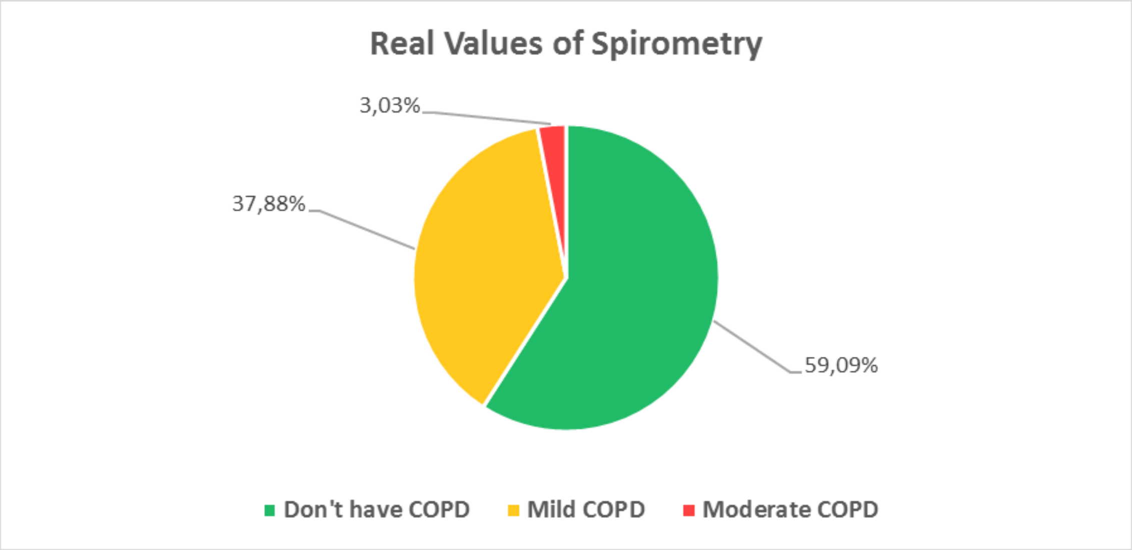


Figure 16 *Circular graphic of real values of spirometry and associated COPD risk.*

The spirometry data indicated that only 3.03% of the population (4 out of 132 individuals) had moderate COPD, 37.88% (50 out of 132 individuals) had mild COPD and 59.09% (76 out of 132 individuals) were not at risk. The COPD prevalence in this study (40.91%) was almost six times higher than the overall rate in China, for women (7%) (Gao and Prasad, 2013). Considering that this condition is an illness that shortens an individual's lifespan and can potentially lead to death (Clayton, 2007), these values are alarming.

4.3.3. The Variables of the Breathing Ratio (FEV1 and FVC)

A stepwise linear regression was made in order to understand what variables had an impact on FEV1 and FVC. These variables are presented in Table 14 below:

Table 14 *Variables that had more impact on FEV1 and FVC.*

FEV1	FVC
Height	Height
Age	Age
Weight	High Blood Pressure
High Blood Pressure	Low Blood Pressure

Height, age, weight and high blood pressure were the variables that had the most impact on FEV1. Height, age and high and low blood pressure were the ones with the most impact on FVC. Age and height constitute the base parameters to calculate the predicted lung volume and the amount of volume expired in one second, in other words the FVC and FEV1. For this reason, it was clear why these two parameters were important. Body weight, which had impact on FEV1, is commonly related with problems in pulmonary function in overweight individuals (Chen, Horne and Dosman, 1993). However, this fact was not relevant to the present study, since none of the individuals was obese. In fact, the body mass index (BMI) of the majority of the subjects was considered normal. High blood pressure was identified as having an impact on both parameters. The association between blood pressure problems and pulmonary function has already been identified (Schnabel et al., 2011). Moreover, one of the causes of pulmonary hypertension is, among others, COPD (Rich, 2012).

4.3.4. Best Linear Model Search

In order to identify which was the best model that could justify the variation of the breathing ratio, a stepwise linear regression was made with all possible variables of the study.

Table 15 Results from stepwise linear regression model.

	AIC
BMI	101.220
Weight	101.235
Age	101.424
Mean Life $PM_{2.5}$ Exposures	101.425
Height	101.533

As shown in table 15, the BMI parameter was the best model to justify the variation in the breathing ratio, since the AIC value was the lowest. BMI is clearly important since both components of the breathing ratio, FEV1 and FVC, are likely to be impacted by body mass and height. These two variables are the basic parameters used to calculate the BMI of each individual (Jones and Nzekwu, 2006).

4.3.5. Variable's Discriminant Analysis

After finding the best model, a linear discriminant analysis was made in order to identify which of the variables of the study might have had higher discriminatory capability on the breathing ratio, FEV1/FVC. The results are based on the Wilks' lambda test, which is used to test which variables contribute significance to a discriminant function. The closer Wilks' lambda is to 0, the more the variable contributes to the discriminant function (Mardia, Bibby and Kent, 1992). The "F" value is related to Wilks' lambda, however the higher the number, the better the result. "F" values >1 are considered significant. The results are presented in Table 16 below:

Table 16 *Linear discriminant analysis of the breathing ratio.*

	Wilks' lambda	F
Stove Used Before Improvement	.989	1.466
Fuel Used Before Improvement	1.000	.039
$PM_{2.5}$ Exposure Before Improvement	.987	1.760
Stove Used After Improvement	1.000	.004
Fuel Used After Improvement	.987	1.678
Age	.981	2.480
Height	1.000	.002
BMI	.973	3.557
Weight	.981	2.477
High Blood Pressure	.999	.133
Low Blood Pressure	.999	.115
Respiratory Tract Infection	1.000	.025

Through the use of the linear discriminant analysis it was possible to identify which variables were more discriminant regarding the breathing ratio. Six variables were found, as showed in Table 16, separated into four levels of discriminatory power. BMI (F=3.557) had the highest discriminant power, followed by age and weight (F=2.480 and F= 2.477), $PM_{2.5}$ exposure before improvement and fuel used after improvement (F=1.760 and F=1.678) and finally the stove used before improvement (F=1.466).

5. Final Conclusion and Perspectives

In the present work, data relating to the exposure of fine atmospheric particulate matter with a dimension below 2.5 micrometers ($PM_{2.5}$) in the interior of the houses in the regions of Xuanwei and Fuyuan was analyzed. Women were chosen as the subjects of this study since, for cultural reasons, they are the ones that are more exposed to this type of contaminants. The primary reason why this demographic group is particularly at risk is due to the fact that all domestic chores, such as cooking, are normally performed by women, making them the ones that spend more time inside the houses and near the stoves, and consequently, near the smoke generated from the combustion of the fuels.

It was concluded that the combination of type of stove and fuel that caused the highest exposure levels of $PM_{2.5}$ belonged to the wood in a fire-pit, with a value of 352 $\mu g/m^3$. This result was particularly alarming, since the annual mean threshold value according to EPA legislation is 15 $\mu g/m^3$. Even after changing the fuel and stove types, that resulted in a reduction of more than 100 $\mu g/m^3$ in the levels of $PM_{2.5}$, the exposure stayed significantly above the recommendations from EPA. Even though the $PM_{2.5}$ exposure values were extremely high, only 3.03% of the population in the data (4 out of 132 individuals) presented moderate COPD, while 37.88% (50 out of 132 individuals) suffered from mild COPD and 59.09% (76 out of 132 individuals) were not at risk. The dispersion of the observed spirometry data was very substantial when compared with the predicted one. This can be justified as a result of natural biological systems associated with humans, showing an intrinsically high diversity due to genetic, social, and cultural aspects. The variables that had the most impact in the individual parameters of the breathing ratio (FEV1 and FVC) were: weight for the FEV1; low blood pressure for the FVC; and height, age and high blood pressure for both. The variable that had the most impact in the breathing ratio was the BMI, while the variables that had the highest discriminant power were the BMI, age, weight, $PM_{2.5}$ exposure before improvement, fuel used after improvement and stove used before improvement. The results of this study showed that there was a significant benefit in the use of smokeless coal, when compared to smoky coal or wood. The primary advantage of the use of this type of fuel is related to the lower $PM_{2.5}$ exposure values, which could potentially reduce the number of breathing problems in the population. However, it might also present other harmful effects similar to the ones caused by smoky coal or wood that are not directly related to $PM_{2.5}$ levels. One way to promote the reduction of these type of medical problems could be the substitution of the use of solid fuels for more efficient fuels, e.g. natural gas or electricity.

In the future, and since the amount of available data was reduced and not ideal, further investigations should be done to support the findings of this work. Particularly, increasing the sample size and producing a better distinction between healthy and unhealthy individuals could yield better and more robust statistical results.

6. References

A Simple Linear Mixed-effects Model. (2010). 1st ed. [ebook] R-Project, p.Chapter 1. Available at: http://lme4.r-forge.r-project.org/book/Ch1.pdf [Accessed 24 Jan. 2017].

Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19(6), pp.716-723.

Balakrishnama, S. and Ganapathiraju, A. (2007). Linear Discriminant Analysis - A Brief Tutorial. 1st ed. [ebook] Mississippi State, Mississippi, EUA. Available at: http://www.music.mcgill.ca/~ich/classes/mumt611_07/classifiers/lda_theory.pdf [Accessed 8 Apr. 2017].

Bellamy, D., Booker, R., Connellan, S. and Halpin, D. (2005). Spirometry in practice: A practical guide to using spirometry in primary care. 2nd ed. [ebook] Available at: https://www.brit-thoracic.org.uk/document-library/delivery-of-respiratory-care/spirometry/spirometry-in-practice-a-practical-guide-(2005)/ [Accessed 15 Feb. 2017].

Benoit, K. (2011). Linear regression models with logarithmic transformations. London School of Economics, London.

Boundless. (2016). The Work of Breathing. [online] Available at: https://www.boundless.com/biology/textbooks/boundless-biology-textbook/the-respiratory-system-39/breathing-221/the-work-of-breathing-839-12084/ [Accessed 30 Mar. 2017].

Caruso, M. S. F., & Alaburda, J. (2008). Hidrocarbonetos policíclicos aromáticos-benzo (a) pireno: uma revisão. Revista do Instituto Adolfo Lutz (Impresso), 67(1), 1-27.

Chafe, Z., Brauer, M., Héroux, M. E., Klimont, Z., Lanki, T., Salonen, R. O., & Smith, K. R. (2015). Residential heating with wood and coal: health impacts and policy options in Europe and North America.

Ccohs.ca. (2017). Environmental Tobacco Smoke (ETS): General Information and Health Effects : OSH Answers. [online] Available at: http://www.ccohs.ca/oshanswers/psychosocial/ets_health.html [Accessed 26 Jul. 2017].

Chambers, J., Becker, R. and Wilks, A. (1988). The new S language. 1st ed.

Chapman, R., Mumford, J., Harris, D., He, X., Jiang, W. and Yang, R. (1990). The epidemiology of lung cancer in Xuan Wei, China; Current progress, issues, and research strategies. Health Policy, 14(2), p.155.

Chen, P. (2000). Study on integrated classification system for Chinese coal. Fuel Processing Technology, 62(2-3), pp.77-87.

Chen, Y., Horne, S. and Dosman, J. (1993). Body weight and weight gain related to pulmonary function decline in adults: a six year follow up study. Thorax, 48(4), pp.375-380.

Clayton, M. (2007). Clinical practice guidelines for communicating prognosis and end-of-life issues with adults in the advanced stages of a life-limiting illness, and their caregivers. 1st ed. Glebe, N.S.W.: Medical Journal of Australia.

Cohen A, Anderson HR, Ostro B, Pandey KD, Krzyzanowski M, Künzli N, Gutschmidt K, Pope CA, Romieu I, Samet JM, Smith KR (2004) Mortality impacts of urban air pollution. In: Ezzati M, Rodgers A, Lopez AD, Murray CJL, eds. Comparative quantification of health risks: global and regional burden of disease due to selected major risk factors. Geneva: World Health Organization, Vol 2.

Cold, F., Health, E., Disease, H., Management, P., Conditions, S., Problems, S., Disorders, S., Checker, S., Interviews, E., Boards, M., Guide, I., Doctor, F., Medications, M., Identifier, P., Interactions, C., Drugs, C., Pregnant, T., Management, D., Obesity, W., Recipes, F., Exercise, F., Beauty, H., Balance, H., Relationships, S., Care, O., Health, W., Health, M., Well, A., Teens, H., Kids, F., Pregnant, G., Trimester, F., Trimester, S., Trimester, T., Baby, N., Health, C., Vaccines, C., Kids, R., Cats, H., Dogs, H., Risks, T., Better?, W., Alzheimer's, '., Syndrome?, G., PTSD, V., Boards, M., Blogs, E., Center, N. and Health, L. (2017). What Are the Stages of COPD?. [online] WebMD. Available at: http://www.webmd.com/lung/copd/gold-criteria-for-copd#1 [Accessed 9 Dec. 2016].

Copdfoundation.org. (2017). What is COPD? | COPD Symptoms, Signs & Risk Factors | COPD Definition. [online] Available at: http://www.copdfoundation.org/What-is-COPD/Understanding-COPD/What-is-COPD.aspx [Accessed 12 Feb. 2017].

Crawley, Michael J. The R book. John Wiley & Sons, 2012.

Division of Cancer Epidemiology and Genetics - National Cancer Institute. (2017). Director's Corner on Global Health: Our Sustained and Productive Partnership with China. [online] Available at: https://dceg.cancer.gov/news-events/linkage-newsletter/2014-11/research-publications/china-partnership [Accessed 13 Jan. 2017].

Dodge, Y. and Marriott, F. (2006). The Oxford dictionary of statistical terms. 1st ed. Oxford, UK: Oxford University Press.

Downward, G. (2015), Quantification and characterization of household air pollution exposure from the use of solid fuels; Clues to the lung cancer epidemic in Xuanwei and Fuyuan, China, Thesis Utrecht University, Chapter 1, ISBN: 978-90-393-6347-8.

Downward, G., Hu, W., Large, D., Veld, H., Xu, J., Reiss, B., Wu, G., Wei, F., Chapman, R., Rothman, N., Qing, L. and Vermeulen, R. (2014). Heterogeneity in coal composition and implications for lung cancer risk in Xuanwei and Fuyuan counties, China. Environment International, 68, pp.94-104.

Enarson, D., Ait-Khaled, N., Chiang, C., El Sony, A., Enarson, P., Gninafon, M., Xiao-Hong, K. and Slama, K. (2009). Lung health consequences of exposure to smoke from domestic use of solid fuels. 1st ed. Paris: International union against tuberculosis and lung disease.

Encyclopedia Britannica. (2017). anthracite | mineral. [online] Available at: https://www.britannica.com/science/anthracite [Accessed 16 Mar. 2017].

Encyclopedia Britannica. (2017). bituminous coal | coal classification. [online] Available at: https://www.britannica.com/science/bituminous-coal [Accessed 16 Mar. 2017].

Freedman, D. (2009). Statistical models. 1st ed. Cambridge: Cambridge University Press.

Gao, J. and Prasad, N. (2013). Chronic obstructive pulmonary disease in China: the potential role of indacaterol. [online] Doi.org. Available at: http://doi.org/10.3978/j.issn.2072-1439.2013.08.04 [Accessed 17 May 2017].

Gardiner, J., Luo, Z. and Roman, L. (2009). Fixed effects, random effects and GEE: What are the differences?. Statistics in Medicine, 28(2), pp.221-239.

Global Alliance for Clean Cook Stoves. (2017). Global Alliance For Clean Cookstoves. [online] Available at: http://cleancookstoves.org [Accessed 29 May 2017].

Glossary of environment statistics. (1997). 1st ed. New York: United Nations.

Gordon, S., Bruce, N., Grigg, J., Hibberd, P., Kurmi, O., Lam, K., Mortimer, K., Asante, K., Balakrishnan, K., Balmes, J., Bar-Zeev, N., Bates, M., Breysse, P., Buist, S., Chen, Z., Havens, D., Jack, D., Jindal, S., Kan, H., Mehta, S., Moschovis, P., Naeher, L., Patel, A., Perez-Padilla, R., Pope, D., Rylance, J., Semple, S. and Martin, W. (2014). Respiratory risks from household air pollution in low and middle income countries. The Lancet Respiratory Medicine, 2(10), pp.823-860.

He, X., Chen, W., Liu, Z. and Chapman, R. (1991). An Epidemiological Study of Lung Cancer in Xuan Wei County, China: Current Progress. Case-Control Study on Lung Cancer and Cooking Fuel. Environmental Health Perspectives, 94, p.9.

Hocking, R. (1976). A Biometrics Invited Paper. The Analysis and Selection of Variables in Linear Regression. Biometrics, 32(1), p.1.

Hong C (1996) Global burden of diseases from air pollution. Shanghai, Shanghai Medical University.

Hosgood, H., Chapman, R., Shen, M., Blair, A., Chen, E., Zheng, T., Lee, K., He, X. and Lan, Q. (2008). Portable stove use is associated with lower lung cancer mortality risk in lifetime smoky coal users. British Journal of Cancer, 99(11), pp.1934-1939.

Howlader, N., Noone, A. and Krapcho, M. (2017). SEER Cancer Statistics Review, 1975-2009 - Previous Version - SEER Cancer Statistics Review. [online] Seer.cancer.gov. Available at: https://seer.cancer.gov/archive/csr/1975_2009_pops09/ [Accessed 10 February 2017].

Jones, R. and Nzekwu, M. (2006). The Effects of Body Mass Index on Lung Volumes. Chest, 130(3), pp.827-833.

Kasteren, H. and Konz, W. (2009). Particulate Matter: Does it matter?. 1st ed. Tilburg, The Netherlands.

Kendzia, B., Gustavsson, P., Pohlabeln, H., Ahrens, W., Jöckel, K., Olsson, A., Gross, I., Brüske, I., Wichmann, H., Merletti, F., Mirabelli, D., Boffetta, P., Straif, K., Brüning, T. and Pesch, B. (2012). Cigarette smoking and lung cancer – Risk estimates for the major histological types from a pooled analysis of European case-control studies. Das Gesundheitswesen, 72(08/09).

Kim, C., Chapman, R., Hu, W., He, X., Hosgood, H., Liu, L., Lai, H., Chen, W., Silverman, D., Vermeulen, R., Tian, L., Bassig, B., Shen, M., Zhang, Y., Ma, S., Rothman, N. and Lan, Q. (2014). Smoky coal, tobacco smoking, and lung cancer risk in Xuanwei, China. Lung Cancer, 84(1), pp.31-35.

Lan, Q., Chapman, R., Schreinemachers, D., Tian, L. and He, X. (2002). Household Stove Improvement and Risk of Lung Cancer in Xuanwei, China. JNCI Journal of the National Cancer Institute, 94(11), pp.826-835.

Lan, Q., He, X., Shen, M., Tian, L., Liu, L., Lai, H., Chen, W., Berndt, S., Hosgood, H., Lee, K., Zheng, T., Blair, A. and Chapman, R. (2008). Variation in lung cancer risk by smoky coal subtype in Xuanwei, China. International Journal of Cancer, 123(9), pp.2164-2169.

Liang, C., Quan, N., Cao, S., He, X. and Ma, F. (1988). Natural inhalation exposure to coal smoke and wood induces lung cancer in mice and rats. Biomed Environmental Science.

Mardia, K., Bibby, J. and Kent, J. (1992). Multivariate analysis. 1st ed. London u.a.: Acad. Pr.

Mumford, J., Chapman, R., Harris, D., He, X., Cao, S., Xian, Y. and Li, X. (1989). Indoor air exposure to coal and wood combustion emissions associated with a high lung cancer rate in Xuan Wei, China. Environment International, 15(1-6), pp.315-320.

Mumford, J., He, X., Chapman, R., Cao, Harris, D., Li, X., Xian, Y., Jiang, W., Xu, C., Chuang, J. and et, a. (1987). Lung cancer and indoor air pollution in Xuan Wei, China. Science, 235(4785), pp.217-220.

[1]Murray CJL, Lopez AD, eds. (1996) In: Global burden of disease and injury series. The global burden of disease: a comprehensive Assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020. Cambridge, Harvard School of Public Health on behalf of the World Health Organization and the World Bank.

[2]Murray, C. and Lopez, A. (1996). The global burden of disease. 1st ed. Cambridge, MA: Published by the Harvard School of Public Health on behalf of the World Health Organization and the World Bank.

Naeher, L., Leaderer, B. and Smith, K. (2000). Particulate Matter and Carbon Monoxide in Highland Guatemala: Indoor and Outdoor Levels from Traditional and Improved Wood Stoves and Gas Stoves. Indoor Air, 10(3), pp.200-205.

National Cancer Institute. (2017). Non-Small Cell Lung Cancer Treatment. [online] Available at: https://www.cancer.gov/types/lung/patient/non-small-cell-lung-treatment-pdq [Accessed 7 Feb. 2017].

Pandey, M., Neupane, R., Gautam, A. and Shrestha, I. (1990). The Effectiveness of Smokeless Stoves in Reducing Indoor Air Pollution in a Rural Hill Region of Nepal. Mountain Research and Development, 10(4), p.313.

Parkin, D., Bray, F., Ferlay, J. and Pisani, P. (2005). Global Cancer Statistics, 2002. CA: A Cancer Journal for Clinicians, 55(2), pp.74-108.

Quanjer, P., Stanojevic, S., Cole, T., Baur, X., Hall, G., Culver, B., Enright, P., Hankinson, J., Ip, M., Zheng, J. and Stocks, J. (2012). Multi-ethnic reference values for spirometry for the 3–95-yr age range: the global lung function 2012 equations. European Respiratory Journal, 40(6), pp.1324-1343.

Rengarajan, T., Rajendran, P., Nandakumar, N., Lokeshkumar, B., Rajendran, P. and Nishigaki, I. (2015). Exposure to polycyclic aromatic hydrocarbons with special focus on cancer. Asian Pacific Journal of Tropical Biomedicine, 5(3), pp.182-189.

Rich, S. (2012). What is pulmonary arterial hypertension?. Pulmonary Circulation, 2(3), pp.271-272.

R-project.org. (2017). R: The R Project for Statistical Computing. [online] Available at: https://www.r-project.org [Accessed 17 Nov. 2016].

Schnabel, E., Karrasch, S., Schulz, H., Gläser, S., Meisinger, C., Heier, M., Peters, A., Wichmann, H., Behr, J., Huber, R. and Heinrich, J. (2011). High blood pressure, antihypertensive medication and lung function in a general adult population. Respiratory Research, 12(1).

Smith KR (1993) Fuel combustion, air pollution exposure, and health: the situation in developing countries. Annual Review of Energy and Environment, 18:529–566.

Smith KR, Mehta S, Feuz M (2004) Indoor smoke from household solid fuels. In: Ezzati M, Rodgers AD, Lopez AD, Murray CJL (eds) Comparative quantification of health risks: Global and regional burden of disease due to selected major risk factors, Geneva: World Health Organization, Vol 2.

Spirometry.guru. (2017). GOLD. [online] Available at: https://www.spirometry.guru/goldcopd.html [Accessed 26 Jan. 2017].

Stewart, B. (2014). World cancer report. 1st ed. Lyon: IARC Press.

Subramanian, J. and Govindan, R. (2007). Lung Cancer in Never Smokers: A Review. Journal of Clinical Oncology, 25(5), pp.561-570.

Sun, S., Schiller, J. and Gazdar, A. (2007). Lung cancer in never smokers — a different disease. Nature Reviews Cancer, 7(10), pp.778-790.

Tian, L., Dai, S., Wang, J., Huang, Y., Ho, S., Zhou, Y., Lucas, D. and Koshland, C. (2008). Nanoquartz in Late Permian C1 coal and the high incidence of female lung cancer in the Pearl River Origin area: a retrospective cohort study. BMC Public Health, 8(1).

Toh, C., Gao, F., Lim, W., Leong, S., Fong, K., Yap, S., Hsu, A., Eng, P., Koong, H., Thirugnanam, A. and Tan, E. (2006). Never-Smokers With Lung Cancer: Epidemiologic Evidence of a Distinct Disease Entity. Journal of Clinical Oncology, 24(15), pp.2245-2251.

Tradingeconomics.com. (2017). Rural population in China. [online] Available at: http://www.tradingeconomics.com/china/rural-population-wb-data.html [Accessed 13 Jan. 2017].

Varma, A. (2009). Applied coal petrology: The role of coal petrology in coal utilization. Isabel Suárez-Ruiz and John C. Crelling (Eds.). Journal of the Geological Society of India, 74(5), pp.652-653.

Webmaster, E. (2017). Global Lung Function Initiative. [online] Lungfunction.org. Available at: http://www.lungfunction.org [Accessed 28 Feb. 2017].

West Cumbria Mining. (2017). What is coking coal? - West Cumbria Mining. [online] Available at: http://www.westcumbriamining.com/what-is-the-plan/what-is-coking-coal/ [Accessed 17 Mar. 2017].

World Health Organization. (2017). Household air pollution and health. [online] Available at: http://www.who.int/mediacentre/factsheets/fs292/en/ [Accessed 5 Feb. 2017]

World Health Organization. (2017). *WHO Global Urban Ambient Air Pollution Database (update 2016)*. [online] Available at:

http://www.who.int/phe/health_topics/outdoorair/databases/cities/en/ [Accessed 29 May 2017].

Xing Y-F, Xu Y-H, Shi M-H, Lian Y-X. The impact of PM2.5 on the human respiratory system. Journal of Thoracic Disease. 2016;8(1):E69-E74. doi:10.3978/j.issn.2072-1439.2016.01.19.

Zhang, L., Lv, J. and Sun, S. (2012). Elements in Lung Tissues of Patients from a High Lung Cancer Incidence Area of China. Biological Trace Element Research, 148(1), pp.7-10.

Zhang, Y. and Cao, F. (2015). Fine particulate matter (PM2.5) in China at a city level. Scientific Reports, 5, p.14884.

7. Appendix

A. Lung Cancer Cases Worldwide

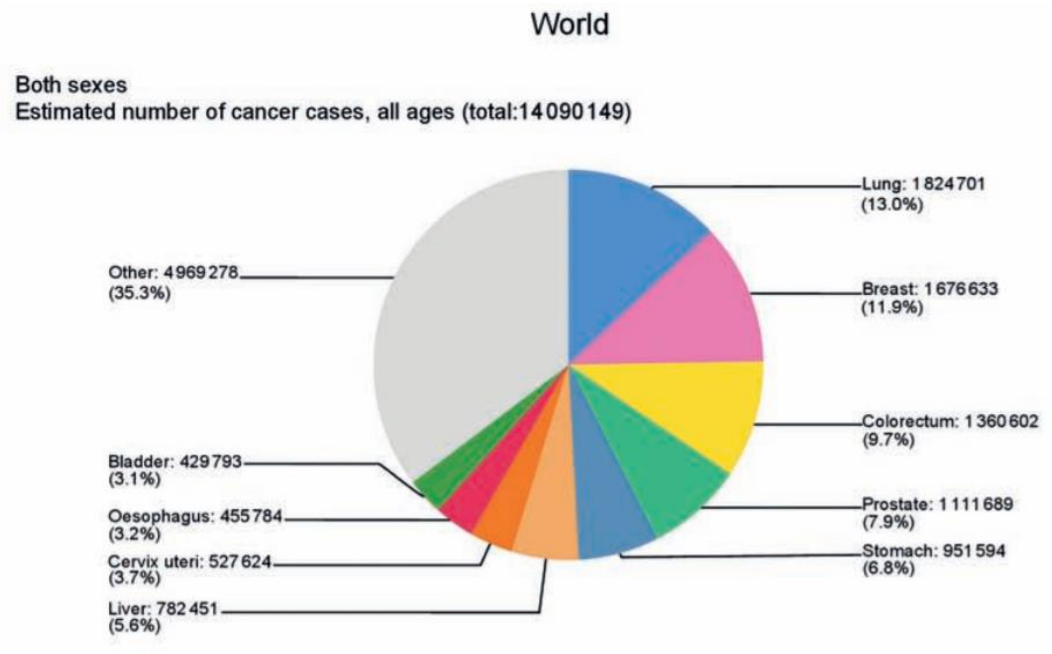


Figure A1 *Estimated world cancer incidence proportions by major sites, in both sexes combined in 2012 (Stewart, 2014).*

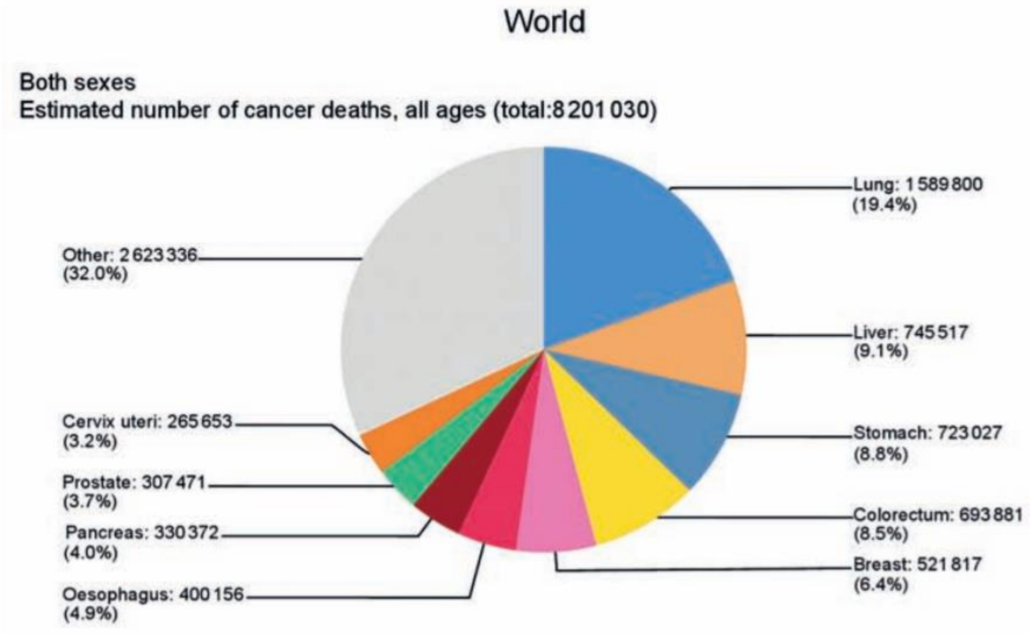


Figure A2 *Estimated world cancer mortality proportions by major sites, in both sexes combined in 2012 (Stewart, 2014).*

B. Coal Types and Subtypes

Coal - Coal is comprised of mineral matter and discrete organic entities, known as macerals, derived from plant debris which has undergone complex changes during coalification (Varma, 2009).

Anthracite coal - Also called hard coal or smokeless coal in the counties of Xuanwei and Fuyuan, the most highly metamorphosed form of coal. It contains more fixed carbon (86% or greater on a dry, ash-free basis) than any other form of coal and the least amount of volatile matter (14% or less on a dry, ash-free basis), and it has calorific values near 35 mega Joules per Kilogram, not much different from the calorific values for most bituminous coal (Encyclopedia Britannica, 2017).

Bituminous coal - Also called soft coal or smoky coal in the counties of Xuanwei and Fuyuan, is the most abundant form of coal, intermediate in rank between subbituminous coal and anthracite according to the coal classification used in the United States and Canada. Bituminous coal is dark brown to black in color and commonly banded, or layered. Because of its relatively high heat value and low (less than 3%) moisture content, its ease of transportation and storage, and its abundance, bituminous coal has the broadest range of commercial uses among the coals (Encyclopedia Britannica, 2017).

Coking coal - Coking coal is an essential ingredient in steel production. It is different to thermal coal which is used to generate power. Coking coal, also known as metallurgic coal, is heated in a coke oven which forces out impurities to produce coke, which is almost pure carbon (West Cumbria Mining, 2017).

Meager Lean Coal – Coal with a lack of desirable qualities, less quality than the refined coal.

C. Variables of the Study

Table A Variables of the study.

Variable ID - (29 variables)	Explanation of the variable
ID	Number assigned to each individual of the study
sex	Sex of the individual
age	Age of the individual at the time the data was collected
ethnic	Ethnicity of each individual
Village_Nr	Village number of each individual
County	County of each individual (That could be Fuyuan or Xuanwei)
Stove_bf	Stove used by each individual before improvement
Fuel_bf	Fuel used by each individual before improvement
PM_Exp_bf	$PM_{2.5}$ exposure that each individual was exposed before improvement
Stove_af	Stove used by each individual after improvement
Fuel_af	Fuel used by each individual after improvement
PM_Exp_af	$PM_{2.5}$ exposure that each individual was exposed after improvement
FVC	Forced Vital Capacity of each individual
FVC_pred	Predicted Forced Vital Capacity of each individual
FEV1	Forced Expiration Volume in 1 second of each individual
FEV1_pred	Predicted Forced Expiration Volume in 1 second of each individual
FEV1/FVC	Breathing ratio
FEV1/FVC_pred	Predicted Breathing ratio
weight	The mean $PM_{2.5}$ exposure value that each individual was exposed in a life-time
BMI	The body index mass of each individual
blood_pressure_h	The high blood pressure of each individual
blood_pressure_l	The low blood pressure of each individual
respiratory_track_infection	If the individuals had some respiratory tract infection at the time of the study

D. Formulas Used in R

Tapply

This function applies a function to each cell of a ragged array, that is to each (non-empty) group of values given by a unique combination of the levels of certain factors) (Chambers, Becker and Wilks, 1988).

Histogram

The generic function "*hist*" computes a histogram of the given data values. If plot = TRUE, the resulting object of class "histogram" is plotted by plot.histogram, before it is returned.

Linear Models

lm is used to fit linear models. It can be used to carry out regression, single stratum analysis of variance and analysis of covariance (although aov may provide a more convenient interface for these) (Chambers, Becker and Wilks, 1988).

Mixed Effect Model

Fit a linear mixed-effects model (LMM) to data, via REML or maximum likelihood.

REML (Restricted Maximum Likelihood)

The REML approach is a particular form of maximum likelihood estimation which does not base estimates on a maximum likelihood fit of all the information, but instead uses a likelihood function calculated from a transformed set of data, so that nuisance parameters have no effect (Dodge and Marriott, 2006).

Akaike Information Criterion (AIC)

AIC is a measure of the relative quality of statistical models for a given set of data, given a collection of models for the data, AIC estimates the quality of each model, relative to each of the other models, where the preferred model is the one with the minimum AIC value. This value rewards goodness of fit and also includes a penalty that is an increasing function of the number of estimated parameters (Akaike, 1974).

E. Study Data

ID	sex	age	height	ethnic	Village_Nr	County	Stove_bf	Fuel_bf	PM_Exp_bf	Stove_af	Fuel_af	PM_Exp_af	Mean_Life_PM_Exposure	FEV1	FEV1_pred	Gold Criteria	FEV1_LLN	FEV1_Z	FEV1_%_pred	FEV1_%_tile	FVC	FVC_pred	FVC_LLN	FVC_Z	FVC_%_pred	FVC_%_tile	FEV1/FVC	FEV1/FVC_pred	FEV1/FVC_LLN	FEV1/FVC_Z	FEV1/FVC_%tile	eNO
102	Female	70,00	155,0	4	1	Fuyuan	Unvented	Smoky	208	PS	Other Coal	191	195,8817	1,850	1,770	A	1,25	0,26	104,5	60,361	2,670	2,198	1,58	1,19	121,4	88,383	0,693	0,808	0,69	-1,63	5,152	24,000
104	Female	43,00	152,0	4	1	Fuyuan	Vented	Wood	222	Vented	Wood	222	222,1427	2,000	2,262	A	1,76	-0,86	88,4	19,479	2,570	2,673	2,10	-0,29	96,1	38,554	0,778	0,847	0,75	-1,20	11,459	7,000
106	Female	54,00	154,0	4	1	Fuyuan	Unvented	Wood	351	Vented	Wood	222	280,8345	2,000	2,103	A	1,58	-0,33	95,1	36,941	2,690	2,544	1,94	0,39	105,8	65,127	0,743	0,828	0,73	-1,41	7,892	5,000
110	Female	42,00	144,0	4	2	Fuyuan	Vented	Wood	222	Vented	Smoky	131	171,4834	1,220	2,034	B	1,58	-2,92	60,0	0,177	1,540	2,379	1,87	-2,76	64,7	0,291	0,792	0,854	0,76	-1,08	13,956	9,000
111	Female	62,00	156,0	4	2	Fuyuan	Unknown	Unknown	-	Vented	Wood	222	222,1427	2,310	1,982	A	1,45	1,05	116,5	85,286	2,650	2,431	1,80	0,55	109,0	71,022	0,872	0,818	0,71	0,93	82,266	NA
114	Female	69,00	147,0	4	2	Fuyuan	Unvented	Wood	351	Vented	Smoky	131	184,7792	1,480	1,602	A	1,14	-0,44	92,4	32,933	2,110	1,972	1,42	0,40	107,0	65,491	0,701	0,814	0,70	-1,60	5,453	30,000
116	Female	61,00	142,0	4	3	Xuanwei	Unvented	Wood	351	Vented	Smoky	131	205,8559	1,400	1,643	A	1,21	-0,92	85,2	17,770	2,350	1,985	1,48	1,14	118,4	87,203	0,596	0,827	0,72	-3,24	0,061	32,000
120	Female	64,00	155,0	4	3	Xuanwei	Vented	Smoky	131	Vented	Smoky	131	131,3782	1,680	1,909	A	1,39	-0,73	88,0	23,274	2,650	2,347	1,73	0,77	112,9	78,043	0,634	0,816	0,71	-2,60	0,472	5,000
122	Female	55,00	150,0	4	4	Fuyuan	Unvented	Wood	351	Vented	Smoky	131	194,2029	1,760	1,969	A	1,48	-0,71	89,4	23,881	3,100	2,376	1,81	1,99	130,5	97,698	0,568	0,829	0,73	-3,73	0,010	13,000
124	Female	59,00	154,0	4	4	Fuyuan	Unvented	Wood	351	Unvented	Wood	351	351,2647	2,270	1,996	A	1,48	0,90	113,8	81,549	3,840	2,432	1,82	3,56	157,9	99,981	0,591	0,822	0,72	-3,30	0,048	21,000
126	Female	57,00	154,0	4	4	Fuyuan	Unvented	Wood	351	Vented	Other Coal	191	243,4333	1,220	2,039	B	1,52	-2,56	59,8	0,524	1,740	2,478	1,87	-2,01	70,2	2,209	0,701	0,825	0,72	-1,95	2,553	49,000
128	Female	71,00	154,0	4	4	Fuyuan	Unvented	Wood	351	Vented	Wood	222	250,8365	1,830	1,723	A	1,21	0,36	106,2	63,965	2,610	2,142	1,53	1,20	121,8	88,542	0,701	0,807	0,69	-1,50	6,725	9,000
131	Female	58,00	158,0	4	5	Xuanwei	Unvented	Wood	351	Vented	Wood	222	263,7244	1,260	2,130	B	1,58	-2,57	59,1	0,503	1,570	2,602	1,96	-2,68	60,3	0,367	0,803	0,821	0,72	-0,32	37,538	NA
132	Female	30,00	150,0	4	5	Xuanwei	Unvented	Wood	351	Vented	Wood	222	313,7777	2,460	2,401	A	1,89	0,19	102,5	57,726	3,190	2,721	2,15	1,31	117,2	90,420	0,771	0,882	0,78	-1,79	3,643	10,000
134	Female	57,00	155,0	4	5	Xuanwei	Unvented	Smoky	208	Vented	Smoky	131	156,3941	2,770	2,068	A	1,54	2,30	134,0	98,920	3,550	2,514	1,90	2,62	141,2	99,558	0,780	0,824	0,72	-0,74	23,081	14,000
141	Female	38,00	162,0	4	6	Xuanwei	Vented	Smoky	131	Vented	Smoky	131	131,3782	2,400	2,696	A	2,11	-0,84	89,0	20,069	3,910	3,169	2,50	1,76	123,4	96,070	0,614	0,853	0,76	-3,50	0,023	8,000
144	Female	67,00	150,0	4	6	Xuanwei	Unknown	Unknown	-	PS	Smoky	177	177,3207	1,600	1,716	A	1,23	-0,40	93,2	34,456	3,020	2,110	1,53	2,44	143,1	99,272	0,530	0,815	0,70	-3,67	0,012	NA
148	Female	69,00	154,0	4	7	Fuyuan	Unvented	Wood	351	Vented	Smoky	131	178,4967	1,930	1,768	A	1,26	0,54	109,1	70,377	2,840	2,191	1,58	1,66	129,6	95,133	0,680	0,810	0,69	-1,84	3,268	19,000
150	Female	17,00	150,0	4	7	Fuyuan	PS	Smoky	177	Unknown	Unknown	-	177,3207	2,420	2,467	A	1,95	-0,15	98,1	43,926	2,760	2,663	2,11	0,28	103,6	61,122	0,877	0,928	0,83	-0,93	17,742	5,000
156	Female	70,00	146,0	4	8	Fuyuan	Unvented	Smoky	207	Vented	Smoky	131	153,9648	2,200	1,559	A	1,10	2,44	141,1	99,262	2,990	1,920	1,38	3,03	155,7	99,877	0,736	0,813	0,70	-1,12	13,214	5,000
158	Female	67,00	148,0	4	8	Fuyuan	Unknown	Unknown	-	Unvented	Wood	351	351,2647	1,840	1,668	A	1,20	0,62	110,3	73,165	2,630	2,047	1,49	1,63	128,5	94,848	0,700	0,816	0,70	-1,68	4,601	27,000
160	Female	72,00	154,0	4	8	Fuyuan	FP	Smoky	221	Vented	Smoky	131	154,7935	1,440	1,714	A	1,19	-0,85	84,7	19,703	2,860	2,118	1,51	1,89	135,0	97,060	0,503	0,806	0,69	-3,71	0,010	14,000
161	Female	75,00	164,0	4	8	Fuyuan	Unvented	Wood	351	Unvented	Wood	351	351,2647	1,920	1,866	A	1,29	0,16	102,9	56,352	2,810	2,360	1,66	1,01	119,1	84,341	0,683	0,797	0,68	-1,55	6,097	16,000
162	Female	77,00	150,0	4	8	Fuyuan	Unvented	Wood	351	Vented	Smoky	131	199,0356	2,240	1,503	A	1,03	2,72	149,0	99,678	3,410	1,885	1,31	4,02	180,9	99,997	0,657	0,803	0,68	-1,89	2,936	8,000
164	Female	40,00	163,0	4	9	Xuanwei	Unvented	Wood	351	Vented	Wood	222	291,4277	2,310	2,689	A	2,10	-1,07	85,9	14,326	3,290	3,184	2,51	0,25	103,3	59,991	0,702	0,847	0,75	-2,37	0,894	21,000
166	Female	66,00	148,0	4	9	Xuanwei	Unvented	Wood	351	Vented	Wood	222	254,905	1,690	1,689	A	1,22	0,00	100,1	50,128	2,060	2,070	1,51	-0,03	99,5	48,900	0,820	0,817	0,71	0,05	51,979	10,000
168	Female	70,00	144,0	4	9	Xuanwei	Unvented	Wood	351	Vented	Wood	222	258,5151	1,270	1,514	A	1,07	-0,92	83,9	17,986	1,920	1,861	1,34	0,18	103,2	57,099	0,661	0,815	0,70	-2,10	1,790	15,000
170	Female	55,00	147,0	4	9	Xuanwei	Unvented	Wood	351	PS	Other Coal	190	236,7185	1,820	1,887	A	1,42	-0,24	96,5	40,621	2,540	2,270	1,73	0,79	111,9	78,533	0,717	0,831	0,73	-1,84	3,265	14,000
171	Female	52,00	156,0	4	9	Xuanwei	Unvented	Wood	351	Unvented	Wood	351	351,2647	1,790	2,205	A	1,67	-1,29	81,2	9,918	2,490	2,662	2,04	-0,45	93,5	32,612	0,719	0,830	0,73	-1,83	3,349	26,000
174	Female	47,00	152,0	4	10	Xuanwei	Vented	Wood	222	Vented	Wood	222	222,1427	2,380	2,184	A	1,68	0,66	109,0	74,408	2,870	2,605	2,03	0,73	110,2	76,842	0,829	0,839	0,74	-0,18	42,727	11,000
176	Female	53,00	144,0	4	10	Xuanwei	Unvented	Wood	351	PS	Smokeless	118	157,1044	2,250	1,842	A	1,39	1,55	122,1	93,903	3,050	2,203	1,69	2,56	138,4	99,474	0,738	0,836	0,74	-1,61	5,339	29,000
178	Female	52,00	156,0	4	10	Xuanwei	Unknown	Unknown	-	Vented	Wood	222	222,1427	1,960	2,205	A	1,67	-0,76	88,9	22,251	2,780	2,662	2,04	0,31	104,4	62,015	0,705	0,830	0,73	-2,04	2,085	113,000
180	Female	84,00	146,0	4	10	Xuanwei	Unvented	Wood	351	Unvented	Wood	351	351,2647	0,730	1,295	B	0,86	-2,12	56,4	1,719	1,280	1,640	1,11	-1,10	78,0	13,611	0,570	0,797	0,66	-2,65	0,400	NA
181	Female	76,00	152,0	4	11	Fuyuan	Unknown	Unknown	-	Unvented	Smokeless	139	138,5637	1,000	1,567	B	1,08	-1,90	63,8	2,856	1,620	1,964	1,37	-0,95	82,5	17,205	0,617	0,803	0,68	-2,37	0,883	24,000
182	Female	77,00	158,0	4	11	Fuyuan	Unknown	Unknown	-	PS	Smokeless	118	118,2723	2,560	1,678	A	1,15	2,93	152,5	99,828	3,290	2,120	1,48	2,78	155,2	99,732	0,778	0,798	0,67	-0,28	39,035	14,000
186	Female	64,00	143,0	4	11	Fuyuan	PS	Smokeless	118	PS	Smokeless	118	118,2723	1,450	1,609	A	1,17	-0,60	90,1	27,332	1,960	1,956	1,44	0,01	100,2	50,513	0,740	0,823	0,71	-1,26	10,373	8,000
188	Female	74,00	142,0	4	11	Fuyuan	Vented	Wood	222	Unvented	Smoky	208	211,1987	1,180	1,393	A	0,97	-0,84	84,7	20,143	1,560	1,723	1,22	-0,52	90,6	30,253	0,756	0,811	0,69	-0,76	22,348	12,000
190	Female	75,00	146,0	4	11	Fuyuan	Unvented	Wood	351	Unvented	Wood	351	351,2647	2,190	1,458	A	1,01	2,85	150,2	99,779	2,970	1,814	1,27	3,27	163,8	99,945	0,737	0,807	0,68	-0,96	16,827	19,000
191	Female	44,00	139,0	4	12	Fuyuan	Unvented	Wood	351	Unvented	Wood	351	227,0021	1,830	1,855	A	1,44	-0,10	98,7	46,001	2,490	2,170	1,70	1,08	114,7	86,078	0,735	0,853	0,76	-1,96	2,498	10,000
192	Female	64,00	154,0	4	12	Fuyuan	Unvented	Wood	351	Unvented	Wood	351	351,2647	2,450	1,883	A	1,37	1,88	130,1	97,028	3,100	2,313	1,70	2,01	134,0	97,765	0,790	0,816	0,71	-0,41	33,997	12,000
194	Female	21,00	153,0	4	12	Fuyuan	Unvented	Wood	351	Unknown	Unknown	-	351,2647	1,610	2,592	B	2,05	-2,91	62,1	0,179	2,370	2,850	2,25	-1,32	83,2	9,388	0,679	0,912	0,81	-3,19	0,072	8,000
201	Female	67,00	149,0	4	18	Fuyuan	Unvented	Wood	351	Vented	Wood	222	256,3221	1,760	1,692	A	1,21	0,24	104,0	59,476	1,860	2,079	1,51	-0,62	89,5	26,676	0,946	0,815	0,70	2,21	98,630	31,000
204	Female	76,00	145,0	4	21	Xuanwei	Vented	Smoky	131	Vented	Smoky	131	131,3782	0,980	1,418	B	0,98	-1,63	69,1	5,133	1,670	1,765	1,23	-0,29	94,6	38,656	0,587	0,807	0,68	-2,74	0,303	8,000
206	Female	72,00	148,0	4	2	Fuyuan	PS	Smoky	177	Vented	Smoky	131	143,3358	1,840	1,562	A	1,10	1,02	117,8	84,534	2,390	1,936	1,38	1,28	123,5	89,910	0,770	0,810	0,69	-0,57	28,284	21,000
208	Female	70,00	144,0	4	2	Fuyuan	Unvented	Wood	351	Vented	Wood	222	190,221	1,380	1,514	A	1,07	-0,51	91,2	30,653	2,450	1,861	1,34	1,75	131,6	95,984	0,563	0,815	0,70	-3,23	0,062	13,000
210	Female	63,00	144,0	4	23	Fuyuan	Unvented	Wood	351	Vented	Smoky	131	196,657	2,060	1,653	A	1,20	1,55	124,6	93,980	2,180	2,008	1,48	0,52	108,6	69,930	0,945	0,823	0,71	2,13	98,323	15,000
302	Female	44,00	154,0	4	13	Xuanwei	Vented	Wood	222	Vented	Wood	222	222,1427	1,760	2,305	B	1,79	-1,73	76,3	4,200	2,730	2,737	2,14	-0,02	99,8	49,267	0,645	0,843	0,75	-3,09	0,101	NA
304	Female	64,00	152,0	4	13	Xuanwei	Unvented	Wood	351	Unvented	Wood	351	351,2647	1,980	1,832	A	1,33	0,50	108,1	69,111	2,430	2,246	1,65	0,49	108,2	68,946	0,815	0,817	0,71	-0,04	48,316	30,000
306	Female	36,00	151,0	4	13	Xuanwei	Vented	Wood	222	Vented	Smoky	131	185,3463	2,780	2,355	A	1,85	1,42	118,0	92,237	3,720	2,724	2,15	2,73	136,6	99,687	0,747	0,865	0,77	-1,93	2,679	12,000
308	Female	55,00	150,0	4	13	Xuanwei	Unvented	Wood	351	Vented	Smoky	131	205,9825	1,360	1,969	B	1,48	-2,03	69,1	2,114	1,950	2,376	1,81	-1,22	82,1	11,047	0,697	0,829	0,73	-2,10	1,805	8,000
310	Female	28,00	142,0	4	13	Xuanwei	Unvented	Wood	351	Unvented	Wood	351	351,2647	2,120	2,156	A	1,70	-0,13	98,3	44,793	2,480	2,409	1,90	0,23	103,0	58,961	0,855	0,893	0,79	-0,67	25,121	18,000
312	Female	42,00	151,0	4	14	Xuanwei	Vented	Wood	222	PS	Smoky	177	199,2105	1,760	2,249	B	1,75	-1,61	78,2	5,328	2,450	2,649	2,08	-0,57	92,5	28,391	0,718	0,849	0,75	-2,17	1,516	7,000
314	Female	66,00	153,0	4	14	Xuanwei	Unknown	Unknown	-	Vented	Smoky	131	131,3782	1,880	1,812	A	1,30	0,22	103,7	58,883	2,570	2,231	1,63	0,89	115,2	81,283	0,732	0,814	0,70	-1,24	10,714	7,000
316	Female	26,00	157,0	4	14	Xuanwei	Vented	Smoky	131	Vented	Smoky	131	131,3782	3,130	2,690	A	2,12	1,31	116,4	90,486	3,890	3,027	2,39	2,14	128,5	98,363	0,805	0,890	0,79	-1,42	7,754	23,000
318	Female	36,00	138,0	4	14	Xuanwei	Unknown	Unknown	-	PS	Smoky	177	177,3207	1,690	1,946	A	1,53	-1,01	86,9	15,581	2,280	2,222	1,76	0,20	102,6	57,962	0,741	0,873	0,77	-2,12	1,710	14,000
320	Female	36,00	160,0	4	15	Xuanwei	PS	Smoky	177	PS	Smoky	177	177,3207	2,740	2,663	A	2,09	0,23	102,9	58,918	4,150	3,105	2,45	2,52	133,6	99,415	0,660	0,859	0,76	-3,01	0,130	11,000
321	Female	72,00	142,0	4	15	Xuanwei	Unknown	Unknown	-	Vented	Wood	222	222,1427	1,240	1,431	A	1,00	-0,75	86,7	22,791	1,770	1,763	1,25	0,02	100,4	50,921	0,701	0,813	0,69	-1,55	6,020	16,000
322	Female	41,00	153,0	4	15	Xuanwei	Vented	Smoky	131	Vented	Smoky	131	131,3782	2,200	2,332	A	1,82	-0,43	94,3	33,382	3,200	2,745	2,16	1,24	116,6	89,267	0,688	0,850	0,75	-2,61	0,457	14,000
328	Female	72,00	160,0	4	16	Xuanwei	Unknown	Unknown	-	PS	Smoky	177	177,3207	1,790	1,843	A	1,29	-0,16	97,1	43,532	2,770	2,309	1,64	1,09	120,0	86,173	0,646	0,803	0,68	-2,13	1,675	11,000
330	Female	45,00	140,0	4	12	Fuyuan	Unvented	Wood	351	PS	Smokeless	118	214,5083	1,320	1,867	B	1,44	-2,11	70,7	1,730	1,860	2,192	1,71	-1,13	84,9	12,823	0,710	0,850	0,75	-2,29	1,102	5,000
331	Female	60,00	154,0	4	16	Xuanwei	Unvented	Wood	351	Vented	Smoky	131	207,4927	2,460	1,973	A	1,45	1,61	124,7	94,584	3,770	2,409	1,80	3,44	156,5	99,971	0,653	0,821	0,72	-2,51	0,600	23,000
332	Female	54,00	151,0	4	16	Xuanwei	PS	Wood	300	Vented	Smoky	131	215,5	1,870	2,017	A	1,52	-0,49	92,7	31,038	2,480	2,433	1,86	0,13	101,9	55,232	0,754	0,830	0,73	-1,27	10,157	15,000
334	Female	33,00	157,0	4	16	Xuanwei	PS	Smoky	177	PS	Smoky	177	177,3207	2,030	2,606	B	2,05	-1,70	77,9	4,457	2,590	3,001	2,37	-1,07	86,3	14,246	0,784	0,870	0,77	-1,45	7,416	13,000
336	Female	45,00	150,0	4	16	Xuanwei	PS	Smoky	177	PS	Smoky	177	177,3207	2,580	2,161	A	1,67	1,45	119,4	92,676	3,060	2,562	2,00	1,41	119,4	92,131	0,843	0,844	0,75	-0,01	49,425	12,000

Table B – Data used to elaborate the study case.

ID	sex	age	height	ethnic	Village_Nr	County	Stove_bf	Fuel_bf	PM_Exp_bf	Stove_af	Fuel_af	PM_Exp_af	Mean_Life_PM_Exposure	FEV1	FEV1_pred	Gold Criteria	FEV1_LLN	FEV1_Z	FEV1_%_pred	FEV1_%_tile	FVC	FVC_pred	FVC_LLN	FVC_Z	FVC_%_pred	FVC_%_tile	FEV1/FVC	FEV1/FVC_pred	FEV1/FVC_LLN	FEV1/FVC_Z	FEV1/FVC_%tile	eNO
338	Female	49,00	153,0	4	17	Fuyuan	Unvented	Wood	351	Vented	Wood	222	273,7915	1,970	2,176	A	1,66	-0,67	90,6	25,134	2,810	2,607	2,02	0,55	107,8	71,008	0,701	0,836	0,74	-2,20	1,377	13,000
340	Female	56,00	156,0	4	17	Fuyuan	Unvented	Wood	351	Unvented	Wood	351	351,2647	2,240	2,118	A	1,58	0,39	105,8	65,015	2,710	2,574	1,95	0,35	105,3	63,631	0,827	0,825	0,72	0,03	51,255	7,000
341	Female	32,00	158,0	4	17	Fuyuan	Unknown	Unknown	-	Vented	Wood	222	222,1427	3,110	2,655	A	2,09	1,36	117,1	91,356	4,690	3,051	2,41	3,97	153,7	99,996	0,663	0,872	0,77	-3,05	0,116	15,000
342	Female	33,00	147,0	4	17	Fuyuan	Unvented	Wood	351	Unvented	Wood	351	351,2647	2,150	2,266	A	1,78	-0,40	94,9	34,461	2,880	2,586	2,04	0,87	111,4	80,714	0,747	0,876	0,78	-2,06	1,957	11,000
346	Female	51,00	142,0	4	18	Fuyuan	Vented	Wood	222	Vented	Wood	222	222,1427	2,510	1,823	A	1,39	2,71	137,7	99,666	2,930	2,169	1,67	2,39	135,1	99,160	0,857	0,840	0,74	0,31	62,253	11,000
348	Female	47,00	159,0	4	18	Fuyuan	Unvented	Wood	351	Unvented	Wood	351	351,2647	2,260	2,403	A	1,85	-0,43	94,0	33,305	2,770	2,884	2,25	-0,29	96,0	38,605	0,816	0,835	0,74	-0,35	36,187	25,000
350	Female	42,00	158,0	4	18	Fuyuan	Vented	Wood	222	Unvented	Wood	351	288,2051	2,350	2,476	A	1,93	-0,38	94,9	35,060	2,950	2,935	2,31	0,04	100,5	51,491	0,797	0,845	0,75	-0,87	19,101	13,000
351	Female	65,00	153,0	4	18	Fuyuan	Unknown	Unknown	-	Unvented	Wood	351	351,2647	1,950	1,835	A	1,33	0,38	106,3	64,879	2,920	2,255	1,65	1,73	129,5	95,776	0,668	0,816	0,70	-2,14	1,614	15,000
352	Female	48,00	143,0	4	18	Fuyuan	Unvented	Wood	351	Vented	Wood	222	277,4807	1,430	1,902	B	1,46	-1,75	75,2	3,981	3,010	2,253	1,75	2,36	133,6	99,083	0,475	0,843	0,75	-4,82	0,000	5,000
354	Female	49,00	147,0	4	19	Fuyuan	Unvented	Wood	351	Unvented	Wood	351	351,2647	2,000	1,999	A	1,53	0,01	100,1	50,203	2,490	2,381	1,84	0,33	104,6	62,770	0,803	0,839	0,74	-0,64	26,139	22,000
356	Female	35,00	156,0	4	19	Fuyuan	Unvented	Wood	351	Unknown	Unknown	-	351,2647	2,690	2,540	A	1,99	0,46	105,9	67,794	3,310	2,942	2,33	0,95	112,5	82,969	0,813	0,865	0,77	-0,91	18,063	8,000
358	Female	54,00	146,0	4	19	Fuyuan	FP	Smokeless	148	PS	Smokeless	118	125,7464	1,490	1,878	B	1,42	-1,38	79,3	8,315	2,080	2,254	1,72	-0,53	92,3	29,871	0,716	0,833	0,73	-1,88	2,975	52,000
361	Female	62,00	148,0	4	19	Fuyuan	FP	Wood	352	Unvented	Smokeless	139	217,129	1,320	1,773	B	1,30	-1,57	74,5	5,869	2,340	2,158	1,60	0,52	108,4	69,815	0,564	0,822	0,71	-3,52	0,021	17,000
362	Female	33,00	151,0	4	20	Fuyuan	PS	Smoky	177	PS	Smoky	177	165,1594	1,600	2,399	B	1,89	-2,54	66,7	0,561	2,330	2,748	2,17	-1,19	84,8	11,731	0,687	0,873	0,77	-2,80	0,257	29,000
364	Female	51,00	140,0	4	20	Fuyuan	Unknown	Unknown	-	FP	Smoky	221	221,3422	1,730	1,769	A	1,35	-0,15	97,8	43,850	2,000	2,100	1,62	-0,34	95,2	36,860	0,865	0,841	0,74	0,44	67,102	13,000
366	Female	50,00	151,0	4	20	Fuyuan	Unknown	Unknown	-	PS	Smoky	177	177,3207	1,650	2,096	B	1,60	-1,48	78,7	6,923	2,300	2,511	1,94	-0,60	91,6	27,421	0,717	0,835	0,74	-1,95	2,556	10,000
368	Female	58,00	144,0	4	20	Fuyuan	FP	Smoky	221	PS	Smoky	177	190,0049	1,630	1,750	A	1,30	-0,44	93,2	32,821	2,480	2,109	1,59	1,13	117,6	86,972	0,657	0,830	0,73	-2,58	0,491	10,000
370	Female	80,00	152,0	4	20	Fuyuan	Unvented	Smoky	208	Unvented	Wood	351	321,1428	1,190	1,485	A	1,00	-1,02	80,1	15,315	1,840	1,877	1,29	-0,10	98,0	45,976	0,647	0,798	0,67	-1,91	2,808	NA
371	Female	39,00	143,0	4	21	Xuanwei	Unknown	Unknown	-	PS	Smoky	177	177,3207	0,620	2,053	C	1,61	-4,99	30,2	0,000	1,510	2,379	1,88	-2,90	63,5	0,187	0,411	0,862	0,76	-5,19	0,000	15,000
372	Female	69,00	155,0	4	21	Xuanwei	Unvented	Wood	351	Vented	Smoky	177	181,6379	2,900	1,793	A	1,27	3,74	161,7	99,991	3,360	2,223	1,60	2,82	151,1	99,759	0,863	0,809	0,69	0,85	80,257	14,000
374	Female	66,00	151,0	4	21	Xuanwei	Unknown	Unknown	-	Vented	Smoky	131	131,3782	1,040	1,763	B	1,27	-2,37	59,0	0,882	1,830	2,166	1,58	-0,93	84,5	17,559	0,568	0,815	0,70	-3,30	0,048	19,000
376	Female	41,00	143,0	4	21	Xuanwei	Vented	Smoky	131	PS	Smoky	177	154,3494	1,610	2,021	B	1,58	-1,52	79,7	6,434	1,970	2,355	1,85	-1,26	83,6	10,395	0,817	0,857	0,76	-0,71	23,971	49,000
378	Female	65,00	159,0	4	21	Xuanwei	Unvented	Wood	351	Vented	Smoky	131	198,0104	2,650	1,991	A	1,44	2,06	133,1	98,010	3,580	2,460	1,80	2,63	145,5	99,578	0,740	0,812	0,70	-1,10	13,513	11,000
380	Female	74,00	155,0	4	22	Xuanwei	Unvented	Wood	351	Unvented	Wood	351	351,2647	1,630	1,678	A	1,17	-0,16	97,2	43,771	2,380	2,101	1,48	0,71	113,3	76,230	0,685	0,803	0,68	-1,61	5,383	27,000
381	Female	60,00	144,0	4	22	Xuanwei	Vented	Smoky	131	Vented	Wood	222	193,8718	1,150	1,711	B	1,26	-2,04	67,2	2,066	1,700	2,069	1,55	-1,15	82,2	12,499	0,676	0,827	0,72	-2,26	1,190	20,000
382	Female	65,00	157,0	4	22	Xuanwei	Unvented	Wood	351	Vented	Wood	222	259,3142	1,600	1,938	A	1,40	-1,04	82,5	14,818	2,690	2,391	1,75	0,74	112,5	77,091	0,595	0,813	0,70	-3,02	0,128	30,000
384	Female	30,00	146,0	4	22	Xuanwei	Vented	Wood	222	PS	Smoky	177	206,2381	2,400	2,267	A	1,78	0,46	105,9	67,877	2,840	2,560	2,02	0,83	111,0	79,793	0,845	0,885	0,78	-0,69	24,378	NA
386	Female	54,00	147,0	4	22	Xuanwei	Unknown	Unknown	-	Vented	Wood	222	222,1427	2,740	1,906	A	1,44	3,07	143,8	99,893	3,120	2,289	1,75	2,39	136,3	99,163	0,878	0,833	0,73	0,84	79,822	17,000
388	Female	35,00	151,0	4	23	Fuyuan	Vented	Smoky	131	Vented	Smoky	131	131,3782	1,310	2,371	B	1,86	-3,35	55,3	0,040	2,260	2,733	2,16	-1,36	82,7	8,770	0,580	0,868	0,77	-3,89	0,005	12,000
390	Female	65,00	154,0	4	23	Fuyuan	Unvented	Wood	351	Vented	Smoky	131	201,3421	1,820	1,861	A	1,34	-0,13	97,8	44,745	2,500	2,289	1,68	0,55	109,2	70,844	0,728	0,815	0,70	-1,31	9,427	12,000
391	Female	38,00	151,0	4	23	Fuyuan	Vented	Wood	222	PS	Smoky	177	202,6049	2,210	2,322	A	1,82	-0,37	95,2	35,520	3,310	2,703	2,13	1,69	122,5	95,461	0,668	0,859	0,76	-2,94	0,166	19,000
392	Female	44,00	151,0	4	23	Fuyuan	Unknown	Unknown	-	PS	Other Coal	191	190,9001	2,160	2,211	A	1,71	-0,17	97,7	43,163	2,930	2,618	2,05	0,88	111,9	81,056	0,737	0,845	0,75	-1,83	3,379	6,000
394	Female	39,00	150,0	4	23	Fuyuan	PS	Smoky	177	PS	Smoky	177	177,3207	1,930	2,272	A	1,78	-1,14	84,9	12,640	2,800	2,651	2,09	0,43	105,6	66,593	0,689	0,857	0,76	-2,65	0,402	25,000
396	Female	22,00	135,0	4	24	Fuyuan	Unvented	Wood	351	Unknown	Unknown	-	351,2647	1,720	1,983	A	1,57	-1,04	86,7	14,844	2,470	2,151	1,70	1,13	114,8	86,983	0,696	0,920	0,82	-3,07	0,106	5,000
401	Female	51,00	150,0	4	24	Fuyuan	Unknown	Unknown	-	Unvented	Wood	351	351,2647	0,750	2,048	C	1,56	-4,17	36,6	0,002	1,400	2,455	1,89	-3,15	57,0	0,082	0,536	0,835	0,74	-4,18	0,001	36,000
406	Female	62,00	151,0	4	25	Xuanwei	Unvented	Wood	351	Vented	Smoky	131	194,2029	2,010	1,850	A	1,35	0,55	108,7	70,747	2,880	2,258	1,67	1,67	127,5	95,248	0,698	0,821	0,71	-1,85	3,186	5,000
408	Female	71,00	149,0	4	25	Xuanwei	Unknown	Unknown	-	Vented	Smoky	131	131,3782	1,300	1,606	A	1,13	-1,07	80,9	14,229	2,230	1,988	1,42	0,68	112,2	75,033	0,583	0,810	0,69	-2,95	0,157	5,000
410	Female	59,00	161,0	4	25	Xuanwei	Unvented	Wood	351	Vented	Smoky	131	219,3328	2,720	2,193	A	1,62	1,58	124,0	94,310	3,340	2,689	2,02	1,52	124,2	93,596	0,814	0,818	0,71	-0,07	47,254	15,000
411	Female	71,00	151,0	4	25	Xuanwei	Unknown	Unknown	-	Vented	Smoky	131	131,3782	2,240	1,652	A	1,16	2,08	135,6	98,125	3,280	2,049	1,46	3,22	160,1	99,936	0,683	0,809	0,69	-1,75	3,988	9,000
412	Female	42,00	150,0	4	25	Xuanwei	Unknown	Unknown	-	Vented	Smoky	131	131,3782	0,610	2,218	D	1,73	-5,07	27,5	0,000	1,480	2,610	2,05	-3,42	56,7	0,032	0,412	0,850	0,75	-5,24	0,000	8,000
414	Female	34,00	153,0	4	26	Xuanwei	Unvented	Wood	351	PS	Smoky	177	286,6569	2,830	2,453	A	1,93	1,22	115,4	88,815	3,520	2,824	2,23	1,86	124,7	96,866	0,804	0,869	0,77	-1,12	13,030	16,000
416	Female	35,00	157,0	4	26	Xuanwei	Vented	Smoky	131	Unvented	Wood	351	198,5657	2,610	2,575	A	2,02	0,11	101,4	54,237	3,750	2,985	2,36	1,93	125,6	97,341	0,696	0,864	0,77	-2,61	0,454	5,000
418	Female	71,00	144,0	4	26	Xuanwei	Vented	Smoky	131	Vented	Smoky	131	149,4087	1,750	1,494	A	1,05	0,99	117,1	83,895	2,410	1,840	1,32	1,69	131,0	95,482	0,726	0,813	0,69	-1,23	10,875	NA
420	Female	34,00	151,0	4	26	Xuanwei	Vented	Smoky	131	PS	Smoky	177	154	1,060	2,385	C	1,87	-4,12	44,4	0,002	3,410	2,741	2,17	1,84	124,4	96,736	0,311	0,870	0,77	-5,40	0,000	6,000
421	Female	36,00	148,0	4	26	Xuanwei	Unknown	Unknown	-	PS	Smoky	177	177,3207	0,940	2,257	C	1,77	-4,29	41,6	0,001	1,350	2,603	2,06	-3,90	51,9	0,005	0,696	0,867	0,77	-2,64	0,415	5,000
422	Female	53,00	153,0	4	27	Fuyuan	Unknown	Unknown	-	Unvented	Smoky	208	207,7426	2,550	2,095	A	1,58	1,52	121,7	93,537	2,870	2,527	1,93	0,92	113,6	82,118	0,889	0,830	0,73	1,09	86,216	11,000
424	Female	67,00	149,0	4	27	Fuyuan	Unvented	Wood	351	PS	Wood	300	313,4416	2,320	1,692	A	1,21	2,26	137,1	98,820	2,920	2,079	1,51	2,30	140,5	98,923	0,795	0,815	0,70	-0,32	37,405	14,000
426	Female	37,00	155,0	4	27	Fuyuan	Unvented	Wood	351	PS	Smoky	177	259,7152	1,810	2,472	B	1,94	-2,03	73,2	2,118	2,470	2,879	2,28	-1,11	85,8	13,410	0,733	0,860	0,76	-2,08	1,890	6,000
501	Female	59,00	155,0	4	27	Fuyuan	Unvented	Wood	351	Unvented	Wood	351	351,2647	1,470	2,023	B	1,50	-1,73	72,7	4,204	2,040	2,468	1,85	-1,13	82,7	12,915	0,721	0,822	0,72	-1,60	5,481	8,000
502	Female	45,00	155,0	4	27	Fuyuan	Unvented	Wood	351	PS	Smoky	177	256,7299	2,750	2,317	A	1,79	1,40	118,7	91,927	4,580	2,759	2,16	4,65	166,0	100,000	0,600	0,841	0,74	-3,59	0,016	12,000
504	Female	71,00	148,0	4	28	Fuyuan	Vented	Wood	222	PS	Wood	300	279,3257	1,450	1,583	A	1,11	-0,48	91,6	31,668	2,860	1,958	1,40	2,49	146,1	99,367	0,507	0,811	0,69	-3,74	0,009	9,000
506	Female	36,00	150,0	4	28	Fuyuan	Vented	Smoky	131	PS	Other Coal	191	153,9	2,680	2,322	A	1,82	1,21	115,4	88,718	3,300	2,683	2,12	1,74	123,0	95,870	0,812	0,865	0,77	-0,94	17,457	8,000
508	Female	57,00	151,0	4	28	Fuyuan	Unvented	Wood	351	PS	Smoky	177	234,3024	1,950	1,956	A	1,46	-0,02	99,7	49,193	3,220	2,370	1,79	2,29	135,9	98,895	0,606	0,826	0,72	-3,22	0,064	11,000
511	Female	37,00	152,0	4	28	Fuyuan	Vented	Wood	222	Vented	Wood	222	222,1427	1,320	2,372	B	1,86	-3,30	55,6	0,048	2,310	2,755	2,18	-1,26	83,9	10,388	0,571	0,861	0,76	-3,97	0,004	19,000
514	Female	79,00	142,0	4	29	Xuanwei	Unvented	Wood	351	Unvented	Wood	351	351,2647	1,570	1,303	A	0,89	1,10	120,5	86,395	2,270	1,627	1,12	1,97	139,5	97,565	0,692	0,805	0,68	-1,46	7,161	11,000
516	Female	67,00	157,0	4	29	Xuanwei	Unvented	Wood	351	Vented	Smoky	131	183,1162	2,270	1,890	A	1,35	1,21	120,1	88,702	2,810	2,340	1,70	1,16	120,1	87,656	0,808	0,811	0,70	-0,05	48,154	10,000
518	Female	67,00	156,0	4	29	Xuanwei	Unvented	Wood	351	Vented	Smoky	131	192,8171	2,400	1,865	A	1,34	1,74	128,7	95,907	3,160	2,306	1,68	2,11	137,0	98,244	0,759	0,811	0,70	-0,79	21,561	31,000
520	Female	44,00	156,0	4	29	Xuanwei	Unvented	Wood	351	PS	Smoky	177	258,4946	2,900	2,369	A	1,84	1,70	122,4	95,499	3,720	2,818	2,21	2,32	132,0	98,994	0,780	0,842	0,75	-1,11	13,353	13,000
521	Female	35,00	150,0	4	30	Xuanwei	Vented	Wood	222	Vented	Wood	222	222,1427	1,800	2,337	B	1,84	-1,76	77,0	3,922	2,180	2,692	2,13	-1,49	81,0	6,783	0,826	0,868	0,77	-0,75	22,570	11,000
522	Female	31,00	144,0	4	30	Xuanwei	Unknown	Unknown	-	PS	Smoky	177	177,3207	1,390	2,191	B	1,72	-2,78	63,4	0,268	1,790	2,478	1,96	-2,19	72,2	1,410	0,777	0,883	0,78	-1,73	4,171	9,000
524	Female	63,00	142,0	4	30	Xuanwei	Vented	Smoky	131	Vented	Smoky	131	131,3782	1,420	1,605	A	1,17	-0,71	88,5	23,976	2,600	1,945	1,44	2,01	133,7	97,776	0,546	0,825	0,72	-3,70	0,011	6,000
526	Female	70,00	157,0	4	30	Xuanwei	Vented	Wood	222	Vented	Wood	222	222,1427	2,810	1,818	A	1,29	3,26	154,5	99,944	3,900	2,263	1,62	3,89	172,3	99,995	0,721	0,807	0,69	-1,25	10,614	17,000
530	Female	48,00	162,0	4	3	Xuanwei	Vented	Smoky	131	Vented	Smoky	131	131,3782	2,730	2,478	A	1,90	0,74	110,2	76,980	3,180	2,988	2,32	0,46	106,4	67,743	0,858	0,832	0,74	0,50	69,304	7,000
531	Female	46,00	147,0	4	3	Xuanwei	PS	Smoky	177	PS	Smoky	177	177,3207	1,710	2,053	A	1,58	-1,21	83,3	11,349	2,380	2,432	1,90	-0,16	97,9	43,813	0,718	0,844	0,75	-2,08	1,874	20,000
534	Female	57,00	152,0	4	4	Fuyuan	Unvented	Wood	351	Vented	Smoky	131	203,41	2,450	1,984	A	1,48	1,58	123,5	94,286	3,500	2,406	1,82	2,88	145,5	99,803	0,700	0,826	0,72	-1,98	2,377	12,000
536	Female	72,00	146,0	4	9	Xuanwei	Unvented	Wood	351	Vented	Wood	222	250,4434	0,940	1,518	B	1,06	-2,08	61,9	1,883	1,330	1,877	1,34	-1,66	70,9	4,814	0,707	0,811	0,69	-1,45	7,414	NA
538	Female	45,00	152,0	4	14	Xuanwei	Unknown	Unknown	-	PS	Smoky	177	177,3207	1,390	2,223	B	1,72	-2,68	62,5	0,368	2,240	2,640	2,06	-1,14	84,8	12,792	0,621	0,843	0,75	-3,37	0,037	11,000
544	Female	34,00	153,0	4	24	Fuyuan	Unvented	Wood	351	Vented	Wood	222	299,6159	2,330	2,453	A	1,93	-0,39	95,0	34,833	2,830	2,824	2,23	0,02	100,2	50,664	0,823	0,869	0,77	-0,81	20,961	9,000

Table B – Data used to elaborate the study case.